



# **Convergence rates of posterior distributions for observations without the iid structure**

**Yang Xing**

**Research Report  
Centre of Biostochastics**

---

Swedish University of  
Agricultural Sciences

Report 2008:11  
ISSN 1651-8543

# Convergence rates of posterior distributions for observations without the iid structure

YANG XING<sup>1</sup>

*Centre of Biostochastics  
Swedish University of Agricultural Sciences  
SE-901 83 Umeå, Sweden*

## Abstract

The classical condition on the existence of uniformly exponentially consistent tests for testing the true density against the complement of its arbitrary neighborhood has been widely adopted in study of asymptotics of Bayesian nonparametric procedures. Because we follow a Bayesian approach, it seems to be more natural to explore alternative and appropriate conditions which incorporate the prior distribution. In this paper we supply a new prior-dependent integration condition to establish general posterior convergence rate theorems for observations which may not be independent and identically distributed. The posterior convergence rates for such observations have recently studied by Ghosal and van der Vaart [5]. We moreover adopt the Hausdorff  $\alpha$ -entropy given by Xing and Ranney [18][16], which is also prior-dependent and smaller than the widely used metric entropies. These lead to extensions of several existing theorems. In particular, we establish a posterior convergence rate theorem for general Markov processes and as its application we improve on the currently known posterior rate of convergence for a nonlinear autoregressive model.

**Keywords:** Density function, Hausdorff entropy, Hellinger metric, infinite-dimensional model, Markov chain, posterior distribution, rate of convergence.

**AMS classification:** 62G20, 62G07, 62F15.

---

<sup>1</sup>E-mail address: yang.xing@sekon.slu.se

# 1 Introduction

The aim of this article is to study the asymptotic behavior of posterior distributions based on observations which are not assumed to be independent and identically distributed. Suppose that  $(\mathfrak{X}^{(n)}, \mathcal{A}^{(n)}, P_\theta^{(n)} : \theta \in \Theta)$ ,  $n = 1, 2, \dots$ , are statistical experiments with observations  $X^{(n)}$ , where the parameter set  $\Theta$  does not depend on the index  $n$ , and suppose that the distributions  $P_\theta^{(n)}$  for all  $\theta \in \Theta$  admit densities  $p_\theta^{(n)}$  relative to a  $\sigma$ -finite measure  $\mu^{(n)}$  on  $\mathfrak{X}^{(n)}$ . Denote by  $\theta_0$  the true parameter generating the observations  $X^{(n)}$ . Assume that  $P_\theta^\infty$  is the infinite product measure  $P_\theta^{(1)} P_\theta^{(2)} \dots P_\theta^{(n)} \dots$  on the product space  $\bigotimes_{n=1}^\infty \mathfrak{X}^{(n)}$ . In the sense that each  $B \subset \mathfrak{X}^{(n)}$  is identified with the subset  $(\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}, \dots, \mathfrak{X}^{(n-1)}, B, \mathfrak{X}^{(n+1)}, \dots)$  of the product space, we have that  $P_\theta^\infty = P_\theta^{(n)}$  holds on  $\mathfrak{X}^{(n)}$  for all  $n$ . In other words,  $P_\theta^\infty$  is the distribution of the sequence  $(X_1, X_2, \dots)$  which makes the observations  $X_n$  independent from  $P_\theta^{(n)}$ . Let  $d_n$  be a semimetric on  $\Theta$ . Note that any semimetric  $d_n(P_{\theta_1}^{(n)}, P_{\theta_2}^{(n)})$  on the space of densities defined on  $\mathfrak{X}^{(n)}$  induces naturally a semimetric  $d_n(\theta_1, \theta_2) = d_n(P_{\theta_1}^{(n)}, P_{\theta_2}^{(n)})$  on  $\Theta$  when the mapping  $\theta \mapsto P_\theta^{(n)}$  is one-to-one which is assumed in the paper. Given a prior  $\Pi_n$  on  $\Theta$ , the posterior distribution  $\Pi_n(\cdot | X^{(n)})$  is a random probability measure given by

$$\Pi_n(B | X^{(n)}) = \frac{\int_B p_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta)}{\int_\Theta p_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta)} = \frac{\int_B R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta)}{\int_\Theta R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta)}$$

for each measurable subset  $B$  in  $\Theta$ , where  $R_\theta^{(n)}(X^{(n)}) = p_\theta^{(n)}(X^{(n)})/p_{\theta_0}^{(n)}(X^{(n)})$  stands for the likelihood ratio. Recall that the posterior distribution  $\Pi_n(\cdot | X^{(n)})$  is said to be convergent almost surely at a rate at least  $\varepsilon_n$  if there exists  $r > 0$  such that  $\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \rightarrow 0$  almost surely as  $n \rightarrow \infty$ . Similarly,  $\Pi_n(\cdot | X^{(n)})$  is said to be convergent in probability at a rate at least  $\varepsilon_n$  if for any sequence  $r_n$  tending to infinity,  $\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r_n\varepsilon_n | X^{(n)}) \rightarrow 0$  in probability as  $n \rightarrow \infty$ . Throughout this paper, almost sure convergence and convergence in probability are understood as to be defined with respect to  $P_{\theta_0}^\infty$ .

Asymptotics of Bayesian nonparametric procedures has been the focus of a considerable amount of research during past three decades. Much works were concerned with the asymptotic behavior of posterior distributions for i.i.d. observations, see, for instance, Barron, Schervish and Wasserman [1], Ghosal, Ghosh and van der Vaart [4], Shen and Wasserman [9] and Walker, Lijoi and Prunster [14]. Recently, Ghosal and van der Vaart [5] proved several types of

posterior convergence rate theorems for non-i.i.d. observations. Their results reply upon the existence of uniformly exponentially consistent tests, combined with the metric entropy condition and the prior concentration rate. Both the existence of uniformly exponentially consistent tests and the metric entropy condition depend on models, but not on priors. Since the posterior depends on the complexity of the model only through the prior, it is therefore of interest to explore alternative conditions which incorporate priors.

In this paper we use an integration condition together with the Hausdorff  $\alpha$ -entropy to study convergence rates of posteriors. The integration condition and the Hausdorff  $\alpha$ -entropy both are prior-dependent. We show that the integration condition is weaker than the existence of uniformly exponentially consistent tests and holds automatically for an interesting class of metrics used to describe rates of convergence. The latter fact leads to an extension of the results for i.i.d. observations in Walker [12][11] and Xing [16], in which construction of such tests is not necessarily required in order to obtain posterior consistency. The integration condition is moreover useful in construction of priors, as shown when we prove that the convergence rates of the pseudoposteriors given by Walker and Hjort [13] do not depend on the metric entropy condition. The Hausdorff  $\alpha$ -entropy condition was introduced in Xing and Ranneby [18][16] and it is weaker than the metric entropy condition. By means of the integration condition and the Hausdorff  $\alpha$ -entropy, we establish general posterior convergence rate theorems both in the almost sure sense and in the in-probability sense. Particularly, we obtain convergence rate theorems of pseudoposteriors and posteriors for independent observations. We also prove a posterior convergence rate theorem for general Markov chains, which is an extension of a result for stationary  $\alpha$ -mixing Markov chains given by Ghosal and van der Vaart ([5], Theorem 5). As applications we improve on the posterior rate of convergence for the nonlinear autoregressive model, see Section 7.4 of Ghosal and van der Vaart [5]. Many authors have studied Bayesian convergence rates for the Gaussian white noise model with a conjugate Gaussian prior (or, equivalently, one has independent normally distributed observations as  $N(\theta_i, 1/n)$ ,  $i = 1, 2, \dots$  and puts a Gaussian prior independently on  $\theta_i$ ,  $i = 1, 2, \dots, n$ ), see for instance Ghosal and van der Vaart [5], Scricciolo [8], Shen and Wasserman [9] and Zhao [20]. Now by our general posterior convergence rate theorem, we extend their results to multi-normally distributed observations which may not be independent.

The paper is organized as follows. In Section 2 we introduce a prior-dependent integration assumption and present several different types of general posterior convergence rate theorems. Section 3 contains applications of our

general results to independent observations and Markov chains. Section 4 contains concrete applications including nonlinear autoregression model, infinite-dimensional normal model and priors based on uniform distributions. The technical proofs are collected in Appendix.

Throughout this paper the notation  $a \lesssim b$  means  $a \leq Cb$  for some positive constant  $C$  which is universal or fixed in the proof. Write  $a \approx b$  if  $a \lesssim b$  and  $b \lesssim a$ . Denote  $Pf^\alpha = \int_{\mathbb{X}} f^\alpha dP$  which is the integral of the nonnegative function  $f$  with power  $\alpha$  relative to the measure  $P$  on  $\mathbb{X}$ .

## 2 General convergence rate theorems

In this section we introduce a new prior-dependent integration condition to study consistency of posterior distributions. The integration condition is shown to be automatically fulfilled by a large number of metrics. Together with the Hausdorff  $\alpha$ -entropy, this integration condition plays a central roll in our versions of general Bayesian convergence rate theorems.

Let us begin with the following assumption given by Ghosal and van der Vaart [5], in which they instead equivalently used a constant multiple of the semimetric  $e_n$ .

**Assumption 1.** *Let  $K$  be a positive constant. Assume that  $\{d_n\}$  and  $\{e_n\}$  are two sequences of semimetrics on  $\Theta$  such that for every  $n$ ,  $\varepsilon > 0$  and  $\theta_1 \in \Theta$  with  $d_n(\theta_1, \theta_0) > \varepsilon$ , there exists a test  $\phi_n$  satisfying*

$$P_{\theta_0}^{(n)} \phi_n \leq e^{-Kn\varepsilon^2} \quad \text{and} \quad \inf_{\theta \in \Theta: e_n(\theta, \theta_1) < \varepsilon} P_{\theta}^{(n)} \phi_n \geq 1 - e^{-Kn\varepsilon^2}.$$

Based on Assumption 1, Ghosal and van der Vaart [5] established a series of general Bayesian convergence rate theorems. Assumption 1 does not depend on the prior distribution. Note that the posterior depends on the complexity of the model only through the prior. As far as the Bayesian approach is concerned, it would be interesting to find some conditions incorporating the prior in study of asymptotic properties. In the following we give such a prior-dependent condition.

Recall that the Hausdorff  $\alpha$ -entropy  $J(\delta, \Theta_1, \alpha, e_n)$  for  $\Theta_1 \subset \Theta$  is the logarithm of the minimal sum of  $\alpha$ -th power of prior masses of balls of  $e_n$ -radius  $\leq \delta$  needed to cover  $\Theta_1$ , see Xing [17] and Xing and Ranney [18] for the details of the Hausdorff  $\alpha$ -entropy. For simplicity of notations, we define the Hausdorff  $\alpha$ -constant  $C(\delta, \Theta_1, \alpha, e_n) := e^{J(\delta, \Theta_1, \alpha, e_n)}$  of any subset  $\Theta_1$  of  $\Theta$ . Observe that  $C(\delta, \Theta_1, \alpha, e_n)$  depends on the prior  $\Pi_n$ . It was proved in Xing

and Ranneby [18] that the inequality

$$\Pi_n(\Theta_1)^\alpha \leq C(\delta, \Theta_1, \alpha, e_n) \leq \Pi_n(\Theta_1)^\alpha N(\delta, \Theta_1, e_n)^{1-\alpha}$$

holds for any  $0 \leq \alpha \leq 1$ , where  $N(\delta, \Theta_1, e_n)$  denotes the minimal number of balls of  $e_n$ -radius  $\leq \delta$  needed to cover  $\Theta_1 \subset \Theta$ . Our prior-dependent integration condition is

**Assumption 2.** *Let  $\{d_n\}$  and  $\{e_n\}$  be two sequences of semimetrics on  $\Theta$ . For some  $\alpha \in (0, 1)$  there exist constants  $K_1 > 0$ ,  $K_2 > 0$  and  $K_3 \geq 0$  such that the inequality*

$$\begin{aligned} & P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1: d_n(\theta, \theta_0) > \varepsilon} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\ & \leq K_1 e^{-K_2 n \varepsilon^2} C(\varepsilon, \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}, \alpha, e_n)^{K_3} \end{aligned}$$

holds for any  $\varepsilon > 0$ ,  $\Theta_1 \subset \Theta$  and for all  $n$  large enough.

We usually take  $K_3 = 1$  but here we let  $K_3 \geq 0$  in order to increase the scope of applicability. It was shown in Xing [17] that Assumption 2 holds when the observations are i.i.d. and  $r e_n = d_n = d$  for some constant  $r > 2$  and some metric  $d$  which is dominated by the Hellinger distance. The integral of Assumption 2 depends on the prior  $\Pi_n$  and hence is trivially equal to zero when  $\Pi_n$  puts zero mass outside of  $\theta_0$ . So Assumption 2 cannot generally imply Assumption 1. In fact, Assumption 2 is weaker than Assumption 1 as shown in the following.

**Proposition 1.** *Assumption 1 implies Assumption 2 for all  $0 < \alpha < 1$ , where one can choose  $K_1 = 2$ ,  $K_2 = (1 - \alpha) K \wedge \alpha K$  and  $K_3 = 1$ .*

We shall use the Hellinger distance  $H(f, g) = \|\sqrt{f} - \sqrt{g}\|_2$  and its modification  $H_*(f, g) = \|(\sqrt{f} - \sqrt{g})(\frac{2}{3}\sqrt{\frac{f}{g}} + \frac{1}{3})^{1/2}\|_2$ , where  $\|h\|_p = (\int_{\mathfrak{X}^{(n)}} |h|^p d\mu^{(n)})^{1/p}$ . The inequalities  $\frac{1}{\sqrt{3}} H(f, g) \leq H_*(f, g) \leq \|f/g\|_\infty^{1/4} H(f, g)$  hold for all densities  $f$  and  $g$ , since  $\|f/g\|_\infty \geq 1$ . The quantity  $H_*$  was used by Xing [16] in computation of prior concentration rates. Denote

$$W_n(\theta_0, \varepsilon) = \left\{ \theta \in \Theta : H_*(p_{\theta_0}^{(n)}, p_\theta^{(n)}) \leq \sqrt{\frac{2}{3}(e^{\frac{3}{2}n\varepsilon^2} - 1)} \right\}.$$

Note that  $W_n(\theta_0, \varepsilon)$  contains the set  $\{\theta \in \Theta : H_*(p_{\theta_0}^{(n)}, p_\theta^{(n)}) \leq \sqrt{n\varepsilon}\}$  because of  $n\varepsilon^2 < \frac{2}{3}(e^{\frac{3}{2}n\varepsilon^2} - 1)$ . The following proposition shows that Assumption

2 holds automatically when  $d_n = e_n = d_n^1$  for some metrics  $d_n^1$  such that  $d_n^1(\theta, \theta_1)^s$  is a convex function of  $\theta$  and

$$d_n^1(\theta_1, \theta_2)^2 \leq -\frac{2}{n} \log \left( 1 - \frac{H(p_{\theta_1}^{(n)}, p_{\theta_2}^{(n)})^2}{2} \right) \quad (1)$$

for all  $n$  and  $\theta_1, \theta_2 \in \Theta$ , where  $s$  is a fixed positive constant. Throughout this paper we let  $d_n^1$  stand for a metric with this property.

**Proposition 2.** *Let  $0 < \delta < 1/2$  and  $0 < \alpha < 1$ . Then the inequality*

$$\begin{aligned} P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1: d_n^1(\theta, \theta_0) > \varepsilon} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\ \leq 2 e^{-\frac{1}{2}(1-\alpha)(1-2\delta)^2 n \varepsilon^2} C(\delta, \varepsilon, \{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon\}, \alpha, d_n^1) \end{aligned}$$

holds for all  $n, \varepsilon > 0$  and  $\Theta_1 \subset \Theta$ .

Another advantage of adoption of Assumption 2 is that it enables us more easily to construct prior distributions  $\Pi_n$  which may receive good posterior convergence rates. Here we present a result which implies that Assumption 2 with  $K_3 = 0$  holds for data-dependent priors  $\Pi_n(d\theta)/p_{\theta}^{(n)}(X^{(n)})^{1-\beta}$  for any given constant  $0 < \beta < 1$ . Data-dependent priors have been studied by Wasserman [15], Walker and Hjort [13] and Xing and Ranney [19].

**Proposition 3.** *The inequality*

$$\begin{aligned} P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1: d_n^1(\theta, \theta_0) > \varepsilon} R_{\theta}^{(n)}(X^{(n)})^\beta \Pi_n(d\theta) \right)^\alpha \\ \leq e^{-((1-\beta) \wedge \beta) \alpha n \varepsilon^2} \Pi_n(\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon)^\alpha \end{aligned}$$

holds for all  $n, 0 < \alpha < 1, 0 < \beta < 1, \varepsilon > 0$  and  $\Theta_1 \subset \Theta$ .

Now we are ready to represent our first main result of this paper.

**Theorem 1.** *Suppose that Assumption 2 holds and that  $\varepsilon_n > 0, n \varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist a constant  $c_1 < K_2$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that*

$$C(j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n\}, \alpha, e_n)^{K_3} \leq e^{c_1 j^2 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n))^\alpha \quad (2)$$

for all sufficiently large integers  $j$  and  $n$ . Then for each  $r$  large enough we have that

$$\Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \longrightarrow 0$$

almost surely as  $n \rightarrow \infty$ . If furthermore there exists  $c_2 > \frac{1}{c_0}$  such that

$$\sum_{n=1}^{\infty} \frac{e^{n\varepsilon_n^2(3+2c_2)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(W_n(\theta_0, \varepsilon_n))} < \infty,$$

then there exists a constant  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \leq e^{-bn\varepsilon_n^2} \quad \text{almost surely}$$

which tends to zero as  $n \rightarrow \infty$ .

Under Assumption 1 and  $\varepsilon_n \gtrsim n^{-\gamma}$  with  $0 < \gamma < 1/2$ , Ghosal and van der Vaart ([5], Theorem 2) proved an almost sure convergence rate theorem and obtained that  $P_{\theta_0}^{(n)} \Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r_n \varepsilon_n | X^{(n)}) = O(\varepsilon_n^2)$  for every  $r_n \rightarrow \infty$ . The upper bound  $\varepsilon_n^2$  is slower than  $e^{-bn\varepsilon_n^2}$  of Theorem 1, and moreover Theorem 1 can be applied to obtain the posterior convergence at the rate  $\varepsilon_n = \sqrt{\log n/n}$ . Note that when  $K_3 = 0$  the inequality (2) follows from  $\Pi_n(W_n(\theta_0, \varepsilon_n)) \geq e^{-\frac{c_1}{\alpha} n\varepsilon_n^2}$ . So Theorem 1 gives that in the special case of  $K_3 = 0$  the concentration rate is precisely equal to the convergence rate. We also mention that in the case that the set  $\Theta$  is convex and  $d_n(\theta, \theta_0)^s$  for some constant  $s > 0$  is a bounded convex function of  $\theta$  in  $\Theta$ , it turns out from Jensen's inequality that the posterior expectation  $\hat{\theta}_n := \int \theta d\Pi_n(\theta | X^{(n)})$  under the assumptions of Theorem 1 yields a point estimator of  $\theta_0$  with the convergence rate at least  $\varepsilon_n$ . Together with Proposition 2, Theorem 1 implies the following direct consequence for the metric  $d_n^1$ .

**Corollary 1.** *Suppose that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist  $0 < \alpha < 1$ ,  $0 < \delta < 1/2$  and  $c_1 < \frac{1}{2}(1 - \alpha)(1 - 2\delta)^2$  such that*

$$C(\delta j \varepsilon_n, \{\theta \in \Theta : j\varepsilon_n < d_n^1(\theta, \theta_0) \leq 2j\varepsilon_n\}, \alpha, d_n^1) \leq e^{c_1 j^2 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n))^\alpha$$

for all sufficiently large integers  $j$  and  $n$ . Then there exists a constant  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d_n^1(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \leq e^{-bn\varepsilon_n^2} \quad \text{almost surely}$$

which tends to zero as  $n \rightarrow \infty$ .

It is also worth pointing out that from Lemma 1 in Xing and Ranney [18] it follows that the inequality (2) can be derived from the following two inequalities:



$$N(j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n\}, e_n)^{K_3(1-\alpha)} \leq e^{c_3 j^2 n \varepsilon_n^2}$$

and

$$\Pi_n(\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n)^{K_3\alpha} \leq e^{c_4 j^2 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n))^\alpha$$

for some constants  $c_3$  and  $c_4$  with  $c_3 + c_4 < K_2$ . Thus, we have the following consequence.

**Corollary 2.** *Suppose that Assumption 2 holds and that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist constants  $c_1, c_2, c_3$  with  $c_1(1-\alpha) + c_2\alpha < K_2$  and  $c_3 > 1/c_0$  and there exists a sequence of subsets  $\Theta_n$  on  $\Theta$  such that for all large  $j$  and  $n$ ,*

- (i)  $N(j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n\}, e_n)^{K_3} \leq e^{c_1 j^2 n \varepsilon_n^2}$ ;
- (ii)  $\Pi_n(\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n)^{K_3} \leq e^{c_2 j^2 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n))$ ;
- (iii)  $\sum_{n=1}^{\infty} \frac{e^{n\varepsilon_n^2(3+2c_3)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(W_n(\theta_0, \varepsilon_n))} < \infty$ .

Then there exists a constant  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \leq e^{-bn\varepsilon_n^2} \quad \text{almost surely}$$

which tends to zero as  $n \rightarrow \infty$ .

Our next theorem gives a slightly weaker version of Theorem 1.

**Theorem 2.** *The following statements are true.*

- (a) *Theorem 1 holds if the inequality (2) is replaced by*

$$C(\varepsilon_n, \Theta_n, \alpha, e_n)^{K_3} \leq e^{c_1 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n))^\alpha \quad \text{for all large } n.$$

- (b) *Corollary 2 holds if both (i) and (ii) are replaced by*

$$N(\varepsilon_n, \Theta_n, e_n)^{K_3} \leq e^{c_1 n \varepsilon_n^2} \quad \text{and} \quad \Pi_n(\Theta_n)^{K_3} \leq e^{c_2 n \varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n)).$$

In order to deal with convergence rates of posterior distributions in the sense of in-probability, following Ghosal and van der Vaart [5], we adopt notations  $V_k(f, g) = \int_{\mathfrak{X}^{(n)}} f |\log(f/g)|^k d\mu^{(n)}$  and  $V_{k,0}(f, g) = \int_{\mathfrak{X}^{(n)}} f |\log(f/g) - K(f, g)|^k d\mu^{(n)}$ , where  $K(f, g) = \int_{\mathfrak{X}^{(n)}} f \log(f/g) d\mu^{(n)}$  is the Kullback-Leibler divergence of densities  $f$  and  $g$ . Denote

$$B_n(\theta_0, \varepsilon; k) = \{\theta \in \Theta : K(p_{\theta_0}^{(n)}, p_{\theta}^{(n)}) \leq n\varepsilon^2, V_{k,0}(p_{\theta_0}^{(n)}, p_{\theta}^{(n)}) \leq n^{k/2}\varepsilon^k\}.$$

Our result in this direction is

**Theorem 3.** *Suppose that Assumption 2 holds and that  $k > 1$ ,  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist a constant  $c_1 < K_2$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that*

$$C(j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n\}, \alpha, e_n)^{K_3} \leq e^{c_1 j^2 n \varepsilon_n^2} \Pi_n(B_n(\theta_0, \varepsilon_n; k))^\alpha \quad (3)$$

for all sufficiently large integers  $j$  and  $n$ . Then for each  $r_n \rightarrow \infty$  we have that

$$\Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r_n \varepsilon_n | X^{(n)}) \longrightarrow 0$$

in probability as  $n \rightarrow \infty$ . If furthermore there exists  $c_2 > 1$  such that  $\frac{e^{c_2 n \varepsilon_n^2} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(B_n(\theta_0, \varepsilon_n; k))} \longrightarrow 0$  as  $n \rightarrow \infty$ , then

$$\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r_n \varepsilon_n | X^{(n)}) \longrightarrow 0$$

in probability as  $n \rightarrow \infty$ .

Similarly, Theorem 3 holds if one replaces the inequality (3) by

$$C(\varepsilon_n, \Theta_n, \alpha, e_n)^{K_3} \leq e^{c_1 n \varepsilon_n^2} \Pi_n(B_n(\theta_0, \varepsilon_n; k))^\alpha \quad \text{for large } n.$$

Moreover, as a consequence of Theorem 3 we obtain the following result which is a slightly stronger version of Theorem 1 in Ghosal and van der Vaart [5].

**Corollary 3.** *Suppose that Assumption 2 holds and that  $k > 1$ ,  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist constants  $c_1, c_2 > 0$  with  $c_1(1 - \alpha) + c_2\alpha < K_2$ ,  $c_3 > 1$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that for all large  $j$  and  $n$ ,*

$$(i) N(j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n\}, e_n)^{K_3} \leq e^{c_1 j^2 n \varepsilon_n^2};$$

$$(ii) \Pi_n(\theta \in \Theta_n : j\varepsilon_n < d_n(\theta, \theta_0) \leq 2j\varepsilon_n)^{K_3} \leq e^{c_2 j^2 n \varepsilon_n^2} \Pi_n(B_n(\theta_0, \varepsilon_n; k));$$

$$(iii) \frac{e^{c_3 n \varepsilon_n^2} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(B_n(\theta_0, \varepsilon_n; k))} \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Then for each  $r_n \rightarrow \infty$  we have that

$$\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r_n \varepsilon_n | X^{(n)}) \longrightarrow 0$$

in probability as  $n \rightarrow \infty$ .

### 3 Some special cases

In this section we apply our general convergence rate theorems to i.n.i.d. observations and Markov processes. For i.n.i.d. observations we establish almost sure convergence rate theorems both on pseudoposterior distributions and on posterior distributions. We derive an almost sure posterior convergence rate theorem for general Markov processes.

#### 3.1 Independent observations

We consider the case that  $X^{(n)}$  is a random vector  $(X_1, X_2, \dots, X_n)$  of independent variables  $X_i$ , where each  $X_i$  is generated from some density  $p_{\theta, i}$  relative to a  $\sigma$ -finite measure  $\mu_i$  on  $(\mathfrak{X}_i, \mathcal{A}_i)$ , and that  $P_\theta^{(n)}$  is the product distribution with the density  $p_\theta^{(n)}(X^{(n)}) = \prod_{i=1}^n p_{\theta, i}(x_i)$  relative to the direct product measure  $\mu^{(n)} = \mu_1 \times \mu_2 \times \dots \times \mu_n$  on  $\mathfrak{X}^{(n)} = \mathfrak{X}_1 \times \mathfrak{X}_2 \times \dots \times \mathfrak{X}_n$ . Assume that  $d_n^0(\theta_1, \theta_2) = (\frac{1}{n} \sum_{i=1}^n H_i(p_{\theta_1, i}, p_{\theta_2, i})^2)^{1/2}$ , where each  $H_i(p_{\theta_1, i}, p_{\theta_2, i}) = (\int (\sqrt{p_{\theta_1, i}} - \sqrt{p_{\theta_2, i}})^2 d\mu_i)^{1/2}$  is the Hellinger distance between  $p_{\theta_1, i}$  and  $p_{\theta_2, i}$  relative to  $\mu_i$  on  $\mathfrak{X}_i$ . It is clear that  $d_n^0$  satisfies the triangle inequality and hence is a metric on  $\Theta$ .

Denote

$$H_{*, i}(p_{\theta_1, i}, p_{\theta_2, i}) = \left( \int (\sqrt{p_{\theta_1, i}} - \sqrt{p_{\theta_2, i}})^2 \left( \frac{2}{3} \sqrt{\frac{p_{\theta_1, i}}{p_{\theta_2, i}}} + \frac{1}{3} \right) d\mu_i \right)^{1/2}.$$

An advantage of adoption of  $H_*$  in computation of concentration rates for independent observations is that we have the following quality

$$\begin{aligned} 1 + \frac{3}{2} H_* \left( \prod_{i=1}^n p_{\theta_1, i}, \prod_{i=1}^n p_{\theta_2, i} \right)^2 &= \prod_{i=1}^n \left( 1 + \frac{3}{2} H_{*, i}(p_{\theta_1, i}, p_{\theta_2, i})^2 \right) \\ &\leq e^{\frac{3}{2} \sum_{i=1}^n H_{*, i}(p_{\theta_1, i}, p_{\theta_2, i})^2}, \end{aligned}$$

which implies that  $W_n(\theta_0, \varepsilon)$  contains the set

$$\overline{W}_n(\theta_0, \varepsilon) := \left\{ \theta \in \Theta : \frac{1}{n} \sum_{i=1}^n H_{*,i}(p_{\theta_0,i}, p_{\theta,i})^2 \leq \varepsilon^2 \right\}.$$

Similarly, we have

$$\begin{aligned} 1 - \frac{1}{2} H \left( \prod_{i=1}^n p_{\theta_1,i}, \prod_{i=1}^n p_{\theta_2,i} \right)^2 &= \prod_{i=1}^n \left( 1 - \frac{1}{2} H_i(p_{\theta_1,i}, p_{\theta_2,i})^2 \right) \\ &\leq e^{-\frac{1}{2} \sum_{i=1}^n H_i(p_{\theta_1,i}, p_{\theta_2,i})^2} = e^{-\frac{1}{2} n d_n^0(\theta_1, \theta_2)^2}, \end{aligned}$$

which implies that the metric  $d_n^0$  satisfies the inequality (1) and hence by the convexity of  $(d_n^0)^2$  one can apply Proposition 2 and Proposition 3 for  $d_n^0$ . Now we are ready to present two results for i.n.i.d. observations by means of  $\overline{W}_n(\theta_0, \varepsilon)$  and  $d_n^0$ .

### 3.1.1 Pseudoposterior convergence rate

Given  $0 < \beta < 1$ , we define a pseudoposterior distribution  $\Pi_{\beta,n}$  based on the prior  $\Pi_n$  by

$$\Pi_{\beta,n}(B \mid X_1, X_2, \dots, X_n) = \frac{\int_B \prod_{i=1}^n p_{\theta,i}(X_i)^\beta \Pi_n(d\theta)}{\int_{\Theta} \prod_{i=1}^n p_{\theta,i}(X_i)^\beta \Pi_n(d\theta)} \quad \text{for each } B \subset \Theta.$$

In other words, we use the data-dependent prior  $\Pi_n(d\theta) / \prod_{i=1}^n p_{\theta,i}(X_i)^{1-\beta}$ .

Wasserman [15] first applied pseudolikelihood function-data-dependent priors in study of asymptotic inference for mixture models. The pseudoposterior  $\Pi_{\beta,n}$  for i.i.d. observations was introduced by Walker and Hjort [13] who proved a Hellinger consistency theorem when  $\beta = 1/2$ . The Hellinger consistency theorem for any  $0 < \beta < 1$  was obtained by Xing and Ranney [19]. Here we study the convergence rates of the pseudoposteriors for i.n.i.d. observations. Using Proposition 3 for  $d_n^0$ , we obtain

**Proposition 4.** *The inequality*

$$\begin{aligned} P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1 : d_n^0(\theta, \theta_0) > \varepsilon} \left( \prod_{i=1}^n \frac{p_{\theta,i}(X_i)}{p_{\theta_0,i}(X_i)} \right)^\beta \Pi_n(d\theta) \right)^\alpha \\ \leq e^{-((1-\beta)\wedge\beta)\alpha n \varepsilon^2} \Pi_n(\theta \in \Theta_1 : d_n^0(\theta, \theta_0) > \varepsilon)^\alpha \end{aligned}$$

holds for all  $n$ ,  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $\varepsilon > 0$  and  $\Theta_1 \subset \Theta$ .

Therefore, we have

**Theorem 4.** *Let  $0 < \beta < 1$ . Suppose that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exists  $c_1 > 0$  such that*

$$\Pi_n(\theta \in \Theta : d_n^0(\theta, \theta_0) > \varepsilon_n) \leq e^{c_1 n \varepsilon_n^2} \Pi_n(\overline{W}_n(\theta_0, \varepsilon_n))$$

for all large  $n$ . Then for each large  $r$ ,

$$\Pi_{\beta, n}(\theta \in \Theta : d_n^0(\theta, \theta_0) \geq r \varepsilon_n | X_1, X_2, \dots, X_n) \longrightarrow 0$$

almost surely as  $n \rightarrow \infty$ .

Since the total mass of  $\Pi_n$  is always equal to one, Theorem 4 implies that the convergence rate  $\varepsilon_n$  of the pseudoposterior distribution  $\Pi_{\beta, n}$  can be completely determined by the concentration condition  $\Pi_n(\overline{W}_n(\theta_0, \varepsilon_n)) \geq e^{c_1 n \varepsilon_n^2}$ . In other words, the convergence rate does not depend on the rate of the metric entropy which describes how large the model is.

### 3.1.2 Posterior convergence rate

By a result of Birgé (see [6], page 491, or [5], Lemma 2) we know that there exist tests satisfying Assumption 1. Based on this fact, Ghosal and van der Vaart ([5], Theorem 4) gave an in-probability convergence rate theorem for i.n.i.d. observations and the metric  $d_n^0$ . Now, together with Proposition 2 and  $\overline{W}_n(\theta_0, \varepsilon) \subset W_n(\theta_0, \varepsilon)$ , Theorem 1 implies the following almost sure assertion.

**Theorem 5.** *Let  $0 < \delta < 1/2$  and  $0 < \alpha < 1$ . Suppose that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist  $c_1 < \frac{1}{2}(1 - \alpha)(1 - 2\delta)^2$ ,  $c_2 > \frac{1}{c_0}$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that*

$$C(\delta j \varepsilon_n, \{\theta \in \Theta_n : j \varepsilon_n < d_n^0(\theta, \theta_0) \leq 2j \varepsilon_n\}, \alpha, d_n^0) \leq e^{c_1 j^2 n \varepsilon_n^2} \Pi_n(\overline{W}_n(\theta_0, \varepsilon_n))^\alpha$$

for all large  $j$ ,  $n$ , and

$$\sum_{n=1}^{\infty} \frac{e^{n \varepsilon_n^2 (3+2c_2)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(\overline{W}_n(\theta_0, \varepsilon_n))} < \infty.$$

Then there exists  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d_n^0(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \leq e^{-bn \varepsilon_n^2} \quad \text{almost surely.}$$

For readers' convenience, we here copy a direct consequence of Theorem 5 for  $\alpha = 1/2$ .

**Corollary 4.** *Let  $0 < \delta < 1/2$ . Suppose that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist  $c_1, c_2, c_3$  with  $c_1 + c_2 < \frac{1}{2}(1 - 2\delta)^2$  and  $c_3 > 1/c_0$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that for all large  $j$  and  $n$ ,*

- (i)  $N(\delta j \varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d_n^0(\theta, \theta_0) \leq 2j\varepsilon_n\}, d_n^0) \leq e^{c_1 j^2 n \varepsilon_n^2}$ ;
- (ii)  $\Pi_n(\theta \in \Theta_n : j\varepsilon_n < d_n^0(\theta, \theta_0) \leq 2j\varepsilon_n) \leq e^{c_2 j^2 n \varepsilon_n^2} \Pi_n(\overline{W}_n(\theta_0, \varepsilon_n))$ ;
- (iii)  $\sum_{n=1}^{\infty} \frac{e^{n\varepsilon_n^2(3+2c_3)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(\overline{W}_n(\theta_0, \varepsilon_n))} < \infty$ .

Then there exists  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d_n^0(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \leq e^{-bn\varepsilon_n^2} \quad \text{almost surely.}$$

### 3.2 Markov chains

Let  $X_0, X_1, \dots$  be a Markov chain with transition density  $p_\theta(y|x)$  and initial density  $q_\theta(x_0)$  with respect to some  $\sigma$ -finite measure  $\mu$  on a measurable space  $(\mathfrak{X}, \mathcal{A})$ . Here we assume that for each  $\theta \in \Theta$  the 2-variable function  $(x, y) \mapsto p_\theta(y|x)$  is measurable. So the joint distribution  $P_\theta^{(n)}$  of  $X_0, X_1, \dots, X_n$  has a density given by  $p_\theta^{(n)}(x^{(n)}) = q_\theta(x_0) \prod_{i=1}^n p_\theta(x_i | x_{i-1})$  relative to the product measure  $\mu(x_0)\mu(x_1) \dots \mu(x_n)$ . We shall adopt the following Hellinger type semimetrics.

$$\begin{aligned} H(p_{\theta_1}(y|x), p_{\theta_2}(y|x)) &= \left( \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{p_{\theta_1}(y|x)} - \sqrt{p_{\theta_2}(y|x)})^2 d\mu(y) d\nu(x) \right)^{1/2}, \\ H(q_{\theta_1}(x), q_{\theta_2}(x)) &= \left( \int_{\mathfrak{X}} (\sqrt{q_{\theta_1}(x)} - \sqrt{q_{\theta_2}(x)})^2 d\mu(x) \right)^{1/2}, \\ H_*(p_{\theta_1}(y|x), p_{\theta_2}(y|x)) &= \left( \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{p_{\theta_1}(y|x)} - \sqrt{p_{\theta_2}(y|x)})^2 \right. \\ &\quad \left. \cdot \left( \frac{2}{3} \sqrt{\frac{p_{\theta_1}(y|x)}{p_{\theta_2}(y|x)}} + \frac{1}{3} \right) d\mu(y) d\nu(x) \right)^{1/2}, \\ H_*(q_{\theta_1}(x), q_{\theta_2}(x)) &= \left( \int_{\mathfrak{X}} (\sqrt{q_{\theta_1}(x)} - \sqrt{q_{\theta_2}(x)})^2 \left( \frac{2}{3} \sqrt{\frac{q_{\theta_1}(x)}{q_{\theta_2}(x)}} + \frac{1}{3} \right) d\mu(x) \right)^{1/2}. \end{aligned}$$

Denote

$$W_n^1(\theta_0, \varepsilon) = \left\{ \theta \in \Theta : H_*(p_{\theta_0}, p_\theta)^2 + \frac{1}{n} H_*(q_{\theta_0}, q_\theta)^2 \leq \varepsilon^2 \right\}.$$

By means of the metric  $d(\theta, \theta_0) := H(p_\theta, p_{\theta_0})$ , Ghosal and van der Vaart ([5], Theorem 5) gave an in-probability posterior convergence rate theorem for stationary  $\alpha$ -mixing Markov chains. Since calculation of the  $\alpha$ -mixing coefficients is generally not easy and many processes are neither mixing nor stationary, it seems worth to develop a posterior convergence rate theorem for Markov chains which may be neither stationary nor  $\alpha$ -mixing. Now we have an almost sure assertion in this direction. Our result is based on the following proposition.

**Proposition 5.** *Suppose that there exist a  $\mu$ -integrable function  $r(y)$  and constants  $a_1 \geq a_0 > 0$  with  $a_1 \geq 1$  such that  $d\nu(y) = r(y)d\mu(y)$  and  $a_0 r(y) \leq p_\theta(y|x) \leq a_1 r(y)$  for all  $\theta \in \Theta$  and  $x, y \in \mathfrak{X}$ . Let  $0 < \delta < \frac{\sqrt{a_0}}{2\sqrt{a_1}}$  and  $0 < \alpha < \frac{1}{2}$ . Then the inequality*

$$\begin{aligned} & P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1: d(\theta, \theta_0) > \varepsilon} \frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{i=1}^n \frac{p_\theta(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} \Pi_n(d\theta) \right)^\alpha \\ & \leq 2 e^{-(\frac{1}{2}-\alpha)(\frac{\sqrt{a_0}}{2}-\sqrt{a_1}\delta)^2 n \varepsilon^2} C(\delta, \varepsilon, \{\theta \in \Theta_1 : d(\theta, \theta_0) > \varepsilon\}, \alpha, d) \end{aligned}$$

holds for all  $n, \varepsilon > 0$  and  $\Theta_1 \subset \Theta$ , where  $d(\theta, \theta_0) = H(p_\theta, p_{\theta_0})$ .

Therefore we have

**Theorem 6.** *Suppose that all assumptions of Proposition 5 hold and suppose that  $\varepsilon_n > 0$ ,  $n \varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist  $c_1 < (\frac{1}{2} - \alpha)(\frac{\sqrt{a_0}}{2} - \sqrt{a_1}\delta)^2$ ,  $c_2 > \frac{1}{c_0}$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that*

$$C(\delta j \varepsilon_n, \{\theta \in \Theta_n : j \varepsilon_n < d(\theta, \theta_0) \leq 2j \varepsilon_n\}, \alpha, d) \leq e^{c_1 j^2 n \varepsilon_n^2} \Pi_n(W_n^1(\theta_0, \varepsilon_n))^\alpha$$

for all large  $j, n$ , and

$$\sum_{n=1}^{\infty} \frac{e^{n \varepsilon_n^2 (3a_1 + 4c_2)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(W_n^1(\theta_0, \varepsilon_n))} < \infty.$$

Then there exists  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d(\theta, \theta_0) \geq r \varepsilon_n | X_0, X_1, \dots, X_n) \leq e^{-bn \varepsilon_n^2} \quad \text{almost surely.}$$

By choosing  $\delta = \frac{\sqrt{a_0}}{4\sqrt{a_1}}$  and  $\alpha = \frac{1}{4}$  we can easily get

**Corollary 5.** *Suppose that there exist a  $\mu$ -integrable function  $r(y)$  and constants  $a_1 \geq a_0 > 0$  such that  $d\nu(y) = r(y)d\mu(y)$  and  $a_0 r(y) \leq p_\theta(y|x) \leq a_1 r(y)$  for all  $\theta \in \Theta$  and  $x, y \in \mathfrak{X}$ . Suppose that  $\varepsilon_n > 0$ ,  $n\varepsilon_n^2 \geq c_0 \log n$  for all large  $n$  and some fixed constant  $c_0 > 0$ . Suppose that there exist  $c_1, c_2, c_3$  with  $3c_1 + c_2 < a_0/16$  and  $c_3 > 1/c_0$  and a sequence of subsets  $\Theta_n$  on  $\Theta$  such that for all large  $j$  and  $n$ ,*

- (i)  $N\left(\frac{\sqrt{a_0}}{4\sqrt{a_1}}j\varepsilon_n, \{\theta \in \Theta_n : j\varepsilon_n < d(\theta, \theta_0) \leq 2j\varepsilon_n\}, d\right) \leq e^{c_1 j^2 n \varepsilon_n^2}$ ;
- (ii)  $\Pi_n(\theta \in \Theta_n : j\varepsilon_n < d(\theta, \theta_0) \leq 2j\varepsilon_n) \leq e^{c_2 j^2 n \varepsilon_n^2} \Pi_n(W_n^1(\theta_0, \varepsilon_n))$ ;
- (iii)  $\sum_{n=1}^{\infty} \frac{e^{n\varepsilon_n^2(3a_1+4c_3)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(W_n^1(\theta_0, \varepsilon_n))} < \infty$ .

Then there exists  $b > 0$  such that for each large  $r$  and all large  $n$ ,

$$\Pi_n(\theta \in \Theta : d(\theta, \theta_0) \geq r\varepsilon_n | X_0, X_1, \dots, X_n) \leq e^{-bn\varepsilon_n^2} \quad \text{almost surely.}$$

## 4 Applications

In this section we give three examples of applications of our theorems. By means of Corollary 5, we improve on the posterior rate of convergence for the nonlinear autoregressive model in Ghosal and van der Vaart [5]. Corollary 1 is applied to find the posterior convergence rate for an infinite-dimensional normal model, which extends the known results in Ghosal and van der Vaart [5], Scricciolo [8], Shen and Wasserman [9] and Zhao [20] for the white noise model with a conjugate prior. Finally, we use Corollary 4 to study priors based on uniform distributions, which extends the corresponding result for priors based on discrete distributions in Ghosal and van der Vaart [5].

### 4.1 Nonlinear autoregression

We observe  $X_1, X_2, \dots, X_n$  of a time series  $\{X_t : t \in \mathbb{Z}\}$  given by

$$X_i = f(X_{i-1}) + \xi_i \quad \text{for } i = 1, 2, \dots, n,$$

where  $\xi_1, \xi_2, \dots, \xi_n$  are i.i.d. random variables with the standard normal distribution and the unknown regression function  $f$  is in the space  $\mathcal{F}$  which consists of all functions  $f$  with  $\sup_{x \in \mathbb{R}} |f(x)| \leq M$  for some fixed positive constant  $M$ .



Let  $q_f(x)$  be the density of  $X_0$  relative to the Lebesgue measure  $d\mu$  on  $\mathbb{R}$ . So  $X_0, X_1, \dots$  can be considered as a Markov chain generated by the transition density  $p_f(y|x) = \phi(y - f(x))$  with  $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$  and the initial density  $q_f(x)$ . Since  $\phi(x)$  is a strictly positive continuous function tending to zero as  $x \rightarrow \pm\infty$ , there exist two constants  $0 < a_0 < 1 < a_1$  depending only on  $M$  such that  $a_0\phi(y) \leq p_f(y|x) \leq a_1\phi(y)$  for all  $f \in \mathcal{F}$  and  $-\infty < y, x < \infty$ . Assume that there exists a constant  $N > 0$  such that the set of initial densities of the Markov chain satisfies  $H_*(q_{f_1}, q_{f_2}) \leq N$  for all initial densities  $q_{f_1}$  and  $q_{f_2}$ . For instance, all of the initial densities with  $a_0\phi(x) \leq q_f(x) \leq a_1\phi(x)$  satisfy  $H_*(q_{f_1}, q_{f_2}) \leq \sqrt{2}(a_1/a_0)^{1/4}$  and hence form a set with the requirement. Define a measure  $d\nu = \phi d\mu$  in  $\mathbb{R}$  and a norm  $\|f\|_2 = (\int_{\mathbb{R}} |f|^2 d\nu)^{1/2}$  on  $\mathcal{F}$ . Assume that the true regression function  $f_0 \in \mathcal{F}$  belongs to the Lipschitz continuous space  $Lip_M$ , which consists of all functions  $f$  on  $(-\infty, \infty)$  satisfying  $|f(x) - f(y)| \leq L|x - y|$  for all  $-\infty < x, y < \infty$ , where  $L$  is a fixed positive constant. When the Markov chain is stationary, Ghosal and van der Vaart ([5], Section 7.4) constructed a prior on the regression functions and obtained the in-probability posterior convergence rate  $n^{-1/3}(\log n)^{1/2}$ , which is the minimax rate times the logarithmic factor  $(\log n)^{1/2}$ . In the following we shall apply Corollary 5 to get the posterior convergence rate  $n^{-1/3}(\log n)^{1/6}$  in the almost sure sense for a general Markov chain defined as above.

First, we note that for any  $f \in \mathcal{F}$ ,

$$\begin{aligned} & H_*(p_{f_0}, p_f)^2 + \frac{1}{n} H_*(q_{f_0}, q_f)^2 \\ & \leq \sqrt{\frac{a_1}{a_0}} H(p_{f_0}, p_f)^2 + \frac{N^2}{n} \\ & = \frac{1}{2} \sqrt{\frac{a_1}{a_0}} \int_{-\infty}^{\infty} \left(1 - e^{-\frac{(f(x)-f_0(x))^2}{4}}\right) d\nu(x) + \frac{N^2}{n} \\ & \leq \frac{\|f - f_0\|_2^2}{8} \sqrt{\frac{a_1}{a_0}} + \frac{N^2}{n}, \end{aligned}$$

where the last inequality follows from the elementary inequality  $1 - e^{-t} \leq t$ . Hence for some small constant  $b_1 > 0$  we have that  $W_n^1(f_0, \varepsilon_n) \supset \{f \in \mathcal{F} : \|f - f_0\|_2 \leq b_1 \varepsilon_n\}$  for all large  $n$ . Similarly,  $\|f - f_0\|_2 \approx H(p_f, p_{f_0})$  hold for all  $f \in \mathcal{F}$  with  $\|f - f_0\|_2 \leq 1$ . Hence Corollary 5 works well for the metric  $\|\cdot\|_2$ .

We also need some basic facts on approximation of Lipschitz continuous functions by means of step functions. Given a finite interval  $[-A_n, A_n]$  and a positive integer  $K_n$ , we make the partition  $[-A_n, A_n) = \bigcup_{k=1}^{K_n} I_k$

with  $I_k = [-A_n + \frac{2A_n(k-1)}{K_n}, -A_n + \frac{2A_n k}{K_n})$  for  $k = 1, 2, \dots, K_n$ . Write  $I_0 = \mathbb{R} \setminus [-A_n, A_n)$ . The space of step functions relative to the partition is the set of functions  $h : [-A_n, A_n) \mapsto \mathbb{R}$  such that  $h$  is identically equal to some constant on each  $I_k$  for  $k = 1, 2, \dots, K_n$ , more precisely,  $h(x) = \sum_{k=1}^{K_n} \beta_k 1_{I_k}(x)$  for some  $\beta = (\beta_1, \beta_2, \dots, \beta_{K_n}) \in [-M, M]^{K_n} \subset \mathbb{R}^{K_n}$ , where  $1_{I_k}(x)$  denotes the indicator function of  $I_k$ . Denote by  $f_\beta(x)$  the function on  $(-\infty, \infty)$  which is equal to  $\sum_{k=1}^{K_n} \beta_k 1_{I_k}(x)$  on  $[-A_n, A_n)$  and vanish outside  $[-A_n, A_n)$ . Hence  $f_\beta \in \mathcal{F}$  and  $\|f_{\beta_1} - f_{\beta_2}\|_2 = \|\beta_1 - \beta_2\|_*$ , where  $\|\beta\|_* = (\sum_{k=1}^{K_n} \beta_k^2 (\int_{I_k} d\nu)^2)^{1/2}$ . Let  $\Pi_n$  be the prior on  $\mathcal{F}$  which is induced by the map  $\beta \mapsto f_\beta$  such that all the coordinates  $\beta_k$  of  $\beta$  are chosen to be i.i.d. random variables with the uniform distribution on  $[-M, M]$ . Hence the support  $\mathcal{F}_n$  of  $\Pi_n$  consists of all such functions  $f_\beta$ . Take  $A_n = 2\sqrt{\log(1/\varepsilon_n)} \approx \sqrt{\log n}$  and  $K_n = \lfloor \frac{3LA_n}{b_1\varepsilon_n} \rfloor + 1$  with  $\varepsilon_n = (\frac{\sqrt{\log n}}{n})^{1/3}$ . Then  $K_n \approx (n \log n)^{1/3} \approx n\varepsilon_n^2$ . Write  $\beta_0 = (\beta_{0,1}, \beta_{0,2}, \dots, \beta_{0,K_n})$  for  $\beta_{0,k} = f_0(-A_n + \frac{2A_n k - 1}{K_n})$ . Since  $f_0 \in \mathcal{F} \cap Lip_L$ , we have that  $f_{\beta_0} \in \mathcal{F}$  and  $\sup_{-A_n \leq x < A_n} |f_{\beta_0}(x) - f_0(x)| \leq LA_n/K_n \leq b_1\varepsilon_n/3$ . From the triangle inequality and the inequality  $\int_x^\infty \phi(t)dt \leq \phi(x)/x$  for all  $x > 0$ , it follows that for all  $f_\beta \in \mathcal{F}_n$  and for all large  $n$ ,

$$\begin{aligned} \left| \|f_\beta - f_0\|_2 - \|f_\beta - f_{\beta_0}\|_2 \right| &\leq \|f_{\beta_0} - f_0\|_2 \\ &= \left( \int_{-A_n}^{A_n} |f_0 - f_{\beta_0}|^2 d\nu \right)^{1/2} + \left( \int_{I_0} f_0^2 d\nu \right)^{1/2} \\ &\leq \frac{b_1\varepsilon_n}{3} \left( \int_{-A_n}^{A_n} d\nu \right)^{1/2} + M \left( \frac{\phi(A_n)}{A_n} \right)^{1/2} \\ &\leq \frac{b_1\varepsilon_n}{3} + \frac{M\varepsilon_n}{(2\pi)^{1/4} A_n^{1/2}} \leq \frac{b_1\varepsilon_n}{2}. \end{aligned}$$

Thus for all large  $j$  and  $n$ , we have

$$\begin{aligned} &\frac{\Pi_n(f_\beta \in \mathcal{F}_n : j\varepsilon_n < \|f_\beta - f_0\|_2 \leq 2j\varepsilon_n)}{\Pi_n(W_n^1(\theta_0, \varepsilon_n))} \\ &\leq \frac{\Pi_n(f_\beta \in \mathcal{F}_n : \|f_\beta - f_0\|_2 \leq 2j\varepsilon_n)}{\Pi_n(f_\beta \in \mathcal{F}_n : \|f_\beta - f_0\|_2 \leq b_1\varepsilon_n)} \\ &\leq \frac{\Pi_n(f_\beta \in \mathcal{F}_n : \|f_\beta - f_0\|_2 \leq 3j\varepsilon_n)}{\Pi_n(f_\beta \in \mathcal{F}_n : \|f_\beta - f_{\beta_0}\|_2 \leq \frac{b_1}{2}\varepsilon_n)} \\ &= \frac{\Pi_n(\beta \in [-M, M]^{K_n} : \|\beta - \beta_0\|_* \leq 3j\varepsilon_n)}{\Pi_n(\beta \in [-M, M]^{K_n} : \|\beta - \beta_0\|_* \leq \frac{b_1}{2}\varepsilon_n)}. \end{aligned}$$

Note that the Euclidean volume of the  $K_n$ -dimensional ellipsoid  $\{\beta \in \mathbb{R}^{K_n} : \|\beta - \beta_0\|_* \leq r\}$  is equal to  $r^{K_n}$  times the Euclidean volume of the "unit"

$K_n$ -dimensional ellipsoid  $\{\beta \in \mathbb{R}^{K_n} : \|\beta - \beta_0\|_* \leq 1\}$ . So the last quotient does not exceed  $j^{2K_n} = e^{K_n \log(2j)}$ , which is less than  $e^{c_2 j^{2n} \varepsilon_n^2}$  for any given  $c_2 > 0$  and all large  $j$ . Hence we have obtained condition (ii) of Corollary 5. Similarly, for all large  $j$  and  $n$ , we have

$$\begin{aligned} & N\left(\frac{\sqrt{a_0}}{4\sqrt{a_1}}j\varepsilon_n, \{f_\beta \in \mathcal{F}_n : j\varepsilon_n < \|f_\beta - f_0\|_2 \leq 2j\varepsilon_n\}, \|\cdot\|_2\right) \\ & \leq N\left(\frac{\sqrt{a_0}}{4\sqrt{a_1}}j\varepsilon_n, \{f_\beta \in \mathcal{F}_n : \|f_\beta - f_{\beta_0}\|_2 \leq 3j\varepsilon_n\}, \|\cdot\|_2\right) \\ & \leq N\left(\frac{\sqrt{a_0}}{4\sqrt{a_1}}j\varepsilon_n, \{\beta \in [-M, M]^{K_n} : \|\beta - \beta_0\|_* \leq 3j\varepsilon_n\}, \|\cdot\|_*\right), \end{aligned}$$

which, by Lemma 4.1 in Pollard [7], is less than  $b_2^{K_n} = e^{K_n \log b_2}$  for some constant  $b_2 > 0$ , and therefore condition (i) of Corollary 5 holds for any given  $c_1 > 0$ .

## 4.2 Infinite-dimensional normal model

We observe an infinite-dimensional random vector  $(X_1, X_2, \dots)$ , where the random vector  $X^{(n)} = (X_1, \dots, X_n)$  for each  $n$  is normally distributed according to  $N(\theta_{(n)}, \Sigma_{(n)})$  with density  $p_{\theta_{(n)}}^{(n)}(x^{(n)})$ ,  $\theta_{(n)} = (\theta_1, \dots, \theta_n)$ , and the covariance matrix  $\Sigma_{(n)}$  is known and satisfies

$$\alpha \Sigma_{(n)}^{-1} \alpha^T \approx n \sum_{i=1}^n \alpha_i^2 \tag{a}$$

for all  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$  and for all  $n$ . The parameter space  $\Theta$  consists of all vectors  $\theta = (\theta_1, \theta_2, \dots)$  in  $\mathbb{R}^\infty$  with  $\|\theta\|_2 := \left(\sum_{i=1}^\infty \theta_i^2\right)^{1/2} < \infty$ . In this section we identify  $\theta_{(n)} = (\theta_1, \dots, \theta_n)$  with  $(\theta_1, \dots, \theta_n, 0, 0, \dots)$  and hence the norm  $\|\theta_{(n)}\|_2$  makes sense. Let  $\gamma$  be a positive constant. The true parameter  $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \dots)$  is assumed to satisfy

$$\sum_{i=1}^\infty \theta_{0,i}^2 i^{2\gamma} < \infty. \tag{b}$$

In the special case that  $X_1, X_2, \dots$  are independent random variables and each  $X_i$  is normally distributed with mean  $\theta_i$  and variance  $1/n$ , the Bayesian estimation problem on parameters  $\theta = (\theta_1, \theta_2, \dots)$  has been studied by many authors including Cox [2], Freedman [3], Ghosal and van der Vaart [5], Scricciolo [8], Shen and Wasserman [9] and Zhao [20]. They showed that posteriors

can attain the minimax rate  $n^{-\gamma/(2\gamma+1)}$ . Observe that every white noise model can be described as an infinite-dimensional normal model via an orthonormal basis.

Now we construct a prior such that the posterior attains the optimal rate of convergence in our framework. We put the prior on the parameter  $\theta = (\theta_1, \theta_2, \dots)$  such that  $\theta_{(k)} = (\theta_1, \dots, \theta_k)$  is distributed as  $N(0, \Sigma_k)$  and that  $\theta_{k+1}, \theta_{k+2}, \dots$  are set to be zero, where  $k = \lfloor n^{1/(2\gamma+1)} c \rfloor$  with some positive constant  $c$  which is determined later and the covariance matrix  $\Sigma_k$  is assumed to satisfy

$$\alpha \Sigma_k^{-1} \alpha^T \lesssim k \sum_{i=1}^k \alpha_i^2 i^{2\gamma} \quad (c)$$

for all  $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$  and for all such  $k$ . For instance, the last inequality holds if eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$  of positive definite matrices  $\Sigma_k^{-1}$  satisfy  $\lambda_i \leq k i^{2\gamma}$  for  $i = 1, 2, \dots, k$ , which for independent variables  $X_1, X_2, \dots$  is slightly weaker than the condition (7.8) given in Ghosal and van der Vaart [5]. In the following we shall apply Corollary 1 to show that the corresponding posterior converges at the rate  $\varepsilon_n = n^{-\gamma/(2\gamma+1)}$ .

**Theorem 7.** *Assume that (a), (b) and (c) hold. Let  $k = \lfloor n^{1/(2\gamma+1)} c \rfloor$  and  $\varepsilon_n = n^{-\gamma/(2\gamma+1)}$ . Then there exist constants  $c > 0$  and  $r > 0$  such that*

$$\Pi_n(\theta \in \Theta : \|\theta - \theta_0\|_2 \geq r \varepsilon_n | X^{(n)}) \longrightarrow 0$$

almost surely as  $n \rightarrow \infty$ .

*Proof.* For any  $\alpha_1 = (\theta_{1,1}, \theta_{1,2}, \dots, \theta_{1,n})$  and  $\alpha_2 = (\theta_{2,1}, \theta_{2,2}, \dots, \theta_{2,n})$  we have

$$\begin{aligned} & H(p_{\alpha_1}^{(n)}, p_{\alpha_2}^{(n)})^2 \\ &= 2 - 2 \int_{\mathbb{R}^n} \sqrt{p_{\alpha_1}^{(n)}(x) p_{\alpha_2}^{(n)}(x)} dx \\ &= 2 - \frac{2}{(2\pi)^{n/2} \sqrt{\det \Sigma_{(n)}}} \\ & \quad \cdot \int_{\mathbb{R}^n} \exp\left(-\frac{1}{4}((x - \alpha_1) \Sigma_{(n)}^{-1} (x - \alpha_1)^T + (x - \alpha_2) \Sigma_{(n)}^{-1} (x - \alpha_2)^T)\right) dx, \end{aligned}$$

where  $x = (x_1, x_2, \dots, x_n)$  and

$$\begin{aligned}
& (x - \alpha_1)\Sigma_{(n)}^{-1}(x - \alpha_1)^T + (x - \alpha_2)\Sigma_{(n)}^{-1}(x - \alpha_2)^T \\
&= 2x\Sigma_{(n)}^{-1}x^T - 2(\alpha_1 + \alpha_2)\Sigma_{(n)}^{-1}x^T + \alpha_1\Sigma_{(n)}^{-1}\alpha_1^T + \alpha_2\Sigma_{(n)}^{-1}\alpha_2^T \\
&= 2\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)\Sigma_{(n)}^{-1}\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)^T \\
&\quad - \frac{1}{2}(\alpha_1 + \alpha_2)\Sigma_{(n)}^{-1}(\alpha_1 + \alpha_2)^T + \alpha_1\Sigma_{(n)}^{-1}\alpha_1^T + \alpha_2\Sigma_{(n)}^{-1}\alpha_2^T \\
&= 2\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)\Sigma_{(n)}^{-1}\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)^T + \frac{1}{2}(\alpha_1 - \alpha_2)\Sigma_{(n)}^{-1}(\alpha_1 - \alpha_2)^T \\
&\geq 2\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)\Sigma_{(n)}^{-1}\left(x - \frac{\alpha_1}{2} - \frac{\alpha_2}{2}\right)^T + b_1 n \|\alpha_1 - \alpha_2\|_2^2
\end{aligned}$$

for some positive constant  $b_1$  independent of  $\alpha_1, \alpha_2$ , where the last inequality follows from condition (a). Hence we get

$$H(p_{\alpha_1}^{(n)}, p_{\alpha_2}^{(n)})^2 \geq 2 - 2e^{-\frac{b_1}{4} n \|\alpha_1 - \alpha_2\|_2^2},$$

which implies that the norm  $2^{-1}b_1\|\cdot\|_2$  satisfies the inequality (1). So Corollary 1 can be applied for the metric  $2^{-1}b_1\|\cdot\|_2$  and for constants  $\alpha = 1/2$  and  $\delta = 1/4$ .

It follows from condition (b) that

$$\begin{aligned}
\|\theta_{(k)} - \theta_0\|_2^2 &= \sum_{i=1}^k (\theta_i - \theta_{0,i})^2 + \sum_{i=k+1}^{\infty} \theta_{0,i}^2 \\
&\leq \|\theta_{(k)} - \theta_{0,(k)}\|_2^2 + k^{-2\gamma} \sum_{i=k+1}^{\infty} \theta_{0,i}^2 i^{2\gamma} \\
&= \|\theta_{(k)} - \theta_{0,(k)}\|_2^2 + O(\varepsilon_n^2),
\end{aligned}$$

where  $\theta_{(k)} = (\theta_1, \dots, \theta_k)$  and  $\theta_{0,(k)} = (\theta_{0,1}, \dots, \theta_{0,k})$ . This implies that for each large  $j$ ,

$$\begin{aligned}
& C\left(\frac{1}{4}j\varepsilon_n, \{\theta_{(k)} : j\varepsilon_n < \frac{b_1}{2}\|\theta_{(k)} - \theta_0\|_2 \leq 2j\varepsilon_n\}, \frac{1}{2}, \frac{b_1}{2}\|\cdot\|_2\right) \\
&\leq C\left(\frac{1}{5}j\varepsilon_n, \{\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n\}, \frac{1}{2}, \|\cdot\|_2\right),
\end{aligned}$$

which by Lemma 1 in Xing and Ranney [18] does not exceed

$$\Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n)^{\frac{1}{2}} N\left(\frac{1}{5}j\varepsilon_n, \{\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n\}, \|\cdot\|_2\right)^{\frac{1}{2}}$$

$$\begin{aligned} &\leq \Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n)^{\frac{1}{2}} b_2^k \\ &\leq \Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n)^{\frac{1}{2}} e^{\frac{1}{40}j^2n\varepsilon_n^2} \end{aligned}$$

for some constant  $b_2 > 1$  and all large  $j, n$ , where we have applied Lemma 4.1 in Pollard [7]. It remains to prove that for large  $j$  and  $n$ ,

$$\Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n) \leq e^{\frac{1}{20}j^2n\varepsilon_n^2} \Pi_n(W_n(\theta_0, \varepsilon_n)).$$

By the proof of Lemma 1 in Xing [16] we have

$$\begin{aligned} 1 + \frac{3}{2} H_*(p_{\theta_{0,(n)}}^{(n)}, p_{\theta_{(n)}}^{(n)})^2 &= E_{\theta_{0,(n)}} \sqrt{p_{\theta_{0,(n)}}^{(n)} / p_{\theta_{(n)}}^{(n)}} = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma_{(n)}}} \\ &\int_{\mathbb{R}^n} \exp\left(-\frac{3}{4}(x - \theta_{0,(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T + \frac{1}{4}(x - \theta_{(n)})\Sigma_{(n)}^{-1}(x - \theta_{(n)})^T\right) dx. \end{aligned}$$

Write

$$\begin{aligned} &-\frac{3}{4}(x - \theta_{0,(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T + \frac{1}{4}(x - \theta_{(n)})\Sigma_{(n)}^{-1}(x - \theta_{(n)})^T \\ &= -\frac{1}{2}(x - \theta_{0,(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T + \frac{1}{2}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}x^T \\ &\quad -\frac{1}{4}\theta_{0,(n)}\Sigma_{(n)}^{-1}\theta_{0,(n)}^T + \frac{1}{4}\theta_{(n)}\Sigma_{(n)}^{-1}\theta_{(n)}^T \\ &= -\frac{1}{2}(x - \theta_{0,(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T + \frac{1}{2}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T \\ &\quad + \frac{1}{4}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(\theta_{0,(n)} - \theta_{(n)})^T \\ &= -\frac{1}{2}((x - \theta_{0,(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T - (\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(x - \theta_{0,(n)})^T \\ &\quad + \frac{1}{4}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(\theta_{0,(n)} - \theta_{(n)})^T) + \frac{3}{8}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(\theta_{0,(n)} - \theta_{(n)})^T \\ &= -\frac{1}{2}(x - \frac{3}{2}\theta_{0,(n)} + \frac{1}{2}\theta_{(n)})\Sigma_{(n)}^{-1}(x - \frac{3}{2}\theta_{0,(n)} + \frac{1}{2}\theta_{(n)})^T \\ &\quad + \frac{3}{8}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(\theta_{0,(n)} - \theta_{(n)})^T. \end{aligned}$$

Hence we obtain

$$1 + \frac{3}{2} H_*(p_{\theta_{0,(n)}}^{(n)}, p_{\theta_{(n)}}^{(n)})^2 = \exp\left(\frac{3}{8}(\theta_{0,(n)} - \theta_{(n)})\Sigma_{(n)}^{-1}(\theta_{0,(n)} - \theta_{(n)})^T\right).$$

It then follows from condition (a) that there exists a positive constant  $b_3$  not depending on  $n$  such that

$$1 + \frac{3}{2} H_*(p_{\theta_{0,(n)}}^{(n)}, p_{\theta_{(n)}}^{(n)})^2 \leq e^{\frac{3}{2}b_3 n \|\theta_{0,(n)} - \theta_{(n)}\|_2^2}.$$

The constant  $c$  is now chosen so largely that  $b_3\|\theta_{(k)} - \theta_{0,(k)}\|_2^2 \leq b_3\|\theta_{(k)} - \theta_{0,(k)}\|_2^2 + 2^{-1}\varepsilon_n^2$ . Since the support  $\Pi_n$  is  $\{(\theta_1, \theta_2, \dots) : \theta_l = 0 \text{ for } l \geq k+1\}$ , we get

$$\Pi_n(W_n(\theta_0, \varepsilon_n)) \geq \Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n)$$

and hence

$$\begin{aligned} & \frac{\Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n)}{\Pi_n(W_n(\theta_0, \varepsilon_n))} \\ & \leq \frac{\Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n)}{\Pi_n(\theta_{(k)} : \|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n)} \\ & = \frac{\int_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n} \exp\left(-\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) d\theta_{(k)}}{\int_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n} \exp\left(-\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) d\theta_{(k)}} \\ & \leq \frac{\int_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq 3j\varepsilon_n} d\theta_{(k)}}{\min_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n} \exp\left(-\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) \int_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n} d\theta_{(k)}} \\ & = \max_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n} \exp\left(\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) \frac{(3j)^k}{((2b_3)^{-1/2})^k} \\ & \leq e^{\frac{1}{40}j^2n\varepsilon_n^2} \max_{\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n} \exp\left(\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) \end{aligned}$$

for all large  $j$  and  $n$ . On the other hand, it turns out from condition (c) that there exists  $b_4 > 0$  such that for any  $\theta_{(k)} = (\theta_1, \dots, \theta_k)$  with  $\|\theta_{(k)} - \theta_{0,(k)}\|_2 \leq (2b_3)^{-1/2}\varepsilon_n$ , we have

$$\begin{aligned} \exp\left(\frac{1}{2}\theta_{(k)}\Sigma_k^{-1}\theta_{(k)}^T\right) & \leq \exp\left(b_4k \sum_{i=1}^k \theta_i^2 i^{2\gamma}\right) \\ & \leq \exp\left(2b_4k \sum_{i=1}^k (\theta_i - \theta_{0,i})^2 i^{2\gamma} + 2b_4k \sum_{i=1}^k \theta_{0,i}^2 i^{2\gamma}\right) \\ & \leq \exp\left(2b_4k^{2\gamma+1} \sum_{i=1}^k (\theta_i - \theta_{0,i})^2 + 2b_4k \sum_{i=1}^{\infty} \theta_{0,i}^2 i^{2\gamma}\right) \\ & \leq \exp\left(b_4b_3^{-1}k^{2\gamma+1}\varepsilon_n^2 + 2b_4k \sum_{i=1}^{\infty} \theta_{0,i}^2 i^{2\gamma}\right) \leq e^{\frac{1}{40}j^2n\varepsilon_n^2} \end{aligned}$$

for all large  $j$  and  $n$ , where the second inequality follows from the inequality  $(s+t)^2 \leq 2s^2 + 2t^2$  for all  $s, t \in \mathbb{R}$ . Therefore, we have proved the required inequality and the proof of Theorem 7 is complete.  $\square$

### 4.3 Prior based on uniform distributions

Assume, just as in Section 3.1, that  $(X_1, X_2, \dots, X_n)$  of independent variables  $X_i$  has a density  $\prod_{i=1}^n p_{\theta,i}(x_i)$  relative to the product measure  $\mu_1 \times \mu_2 \times \dots \times \mu_n$  on  $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \dots \times \mathfrak{X}_n$ . We follow the notations of Section 3.1. By means of the componentwise Hellinger upper bracketing numbers for  $\Theta$ , Ghosal and van der Vaart [5] have obtained an in-probability convergence rate theorem for priors based on discrete distributions. Their result can be extended to an almost sure assertion in terms of Theorem 5. In the following we give an almost sure result for priors based on uniform distributions, which gives us an opportunity to adopt the average Hellinger metric  $d_n^0(\theta_1, \theta_2) = \left(\frac{1}{n} \sum_{i=1}^n H_i(p_{\theta_1,i}, p_{\theta_2,i})^2\right)^{1/2}$  instead of the componentwise Hellinger upper bracketing numbers. This also extends a result for i.i.d. observations given by Xing ([16], Section 3.2).

Let  $c > 1$  and let  $\bar{d}_n$  be metrics on  $\Theta$ . Assume that  $\Theta_{c,n}$  for  $n = 1, 2, \dots$  are subsets of  $\Theta$  such that  $\frac{1}{c^2} \sum_{i=1}^n H_{*,i}(p_{\theta_1,i}, p_{\theta_2,i})^2 \leq \bar{d}_n(\theta_1, \theta_2)^2$  for all  $\theta_1, \theta_2 \in \Theta_{c,n}$ . By the definition of  $H_{*,i}$  we have  $d_n^0 \leq \sqrt{3}c \bar{d}_n$  on  $\Theta_{c,n}$ . Note that  $\bar{d}_n$  can be taken as a constant multiple of  $d_n^0$  in the case that  $H_{*,i}(p_{\theta_1,i}, p_{\theta_2,i}) \lesssim H_i(p_{\theta_1,i}, p_{\theta_2,i})$  for all  $\theta_1, \theta_2$  in  $\Theta$  and  $i = 1, 2, \dots, n$ . Given  $\varepsilon_n > 0$ , we assume that  $\{B_1, \dots, B_{K_n}\}$  is a partition of  $\Theta_{c,n}$  such that for each  $B_i$  there exists  $b_i$  in  $\Theta$  with  $B_i \subset \{\theta \in \Theta_{c,n} : \bar{d}_n(b_i, \theta) \leq \varepsilon_n/2c\}$ . Let  $\Pi_n$  be a prior distribution supported on  $\Theta_{c,n}$  such that  $\Pi_n(B_i) = 1/K_n$  for  $i = 1, 2, \dots, K_n$ . Corollary 4 implies the following result.

**Theorem 8.** *Suppose that  $\theta_0 \in \Theta_{c,n}$  for all  $n$  and suppose that  $\log K_n + \log n = O(n\varepsilon_n^2)$  as  $n \rightarrow \infty$ . Then for each large  $r$ ,*

$$\Pi_n(\theta \in \Theta : d_n^0(\theta, \theta_0) \geq r\varepsilon_n | X_1, X_2, \dots, X_n) \rightarrow 0$$

*almost surely as  $n \rightarrow \infty$ .*

*Proof.* Take  $\Theta_n = \Theta_{c,n}$  for all  $n$ . Then condition (iii) of Corollary 4 is trivially fulfilled. For  $\delta = 1/(2\sqrt{3}c^2)$  we have that for any given  $c_1 > 0$  and all large  $j$  and  $n$ ,

$$\begin{aligned} N(\delta j \varepsilon_n, \{\theta \in \Theta_n : j \varepsilon_n < d_n^0(\theta, \theta_0) \leq 2j \varepsilon_n\}, d_n^0) &\leq N\left(\frac{\varepsilon_n}{2\sqrt{3}c^2}, \Theta_n, d_n^0\right) \\ &\leq N\left(\frac{\varepsilon_n}{2c}, \Theta_n, \bar{d}_n\right) \leq K_n \leq e^{c_1 j^2 n \varepsilon_n^2}, \end{aligned}$$



where the last inequality follows from  $\log K_n = O(n\varepsilon_n^2)$ . This implies condition (i) of Corollary 4. To see condition (ii), by  $\theta_0 \in \Theta_{c,n}$  we can take  $b_{i_0} \in \Theta$  such that  $\bar{d}_n(b_{i_0}, \theta_0) \leq \varepsilon_n/2c$ . Then, for all  $\theta \in B_{i_0}$  we have

$$\frac{1}{n} \sum_{i=1}^n H_{*,i}(p_{\theta_0,i}, p_{\theta,i})^2 \leq c^2 \bar{d}_n(\theta_0, \theta)^2 \leq c^2 (\bar{d}_n(\theta_0, b_{i_0}) + \bar{d}_n(b_{i_0}, \theta))^2 \leq \varepsilon_n^2,$$

which implies that  $\bar{W}_n(\theta_0, \varepsilon_n)$  contains the whole set  $B_{i_0}$  and hence  $\Pi_n(\bar{W}_n(\theta_0, \varepsilon_n)) \geq \Pi_n(B_{i_0}) = 1/K_n \geq e^{-c_2 j^2 n \varepsilon_n^2}$  for any given  $c_2 > 0$  and all large  $j$  and  $n$ . So we have verified condition (ii) and the proof of Theorem 8 is complete.  $\square$

*Example* (Nonparametric Poisson regression): Assume that  $U \geq L > 0$  are two given constants. We consider Poisson distributed independent random variables  $X_1, X_2, \dots, X_n$  with parameters  $\theta(z_1), \theta(z_2), \dots, \theta(z_n)$ , where  $\theta : \mathbb{R} \rightarrow [L, U]$  is an unknown increasing link function and  $z_1, z_2, \dots, z_n$  are one-dimensional covariates. The joint mass function of  $(X_1, X_2, \dots, X_n)$  is given by  $\prod_{i=1}^n p_{\theta,i}(x_i)$  with  $p_{\theta,i}(x_i) = e^{-\theta(z_i)} \frac{\theta(z_i)^{x_i}}{x_i!}$ . For  $a, b \in [L, U]$  we have

$$\begin{aligned} & \sum_{x=0}^{\infty} \left( \sqrt{e^{-a} \frac{a^x}{x!}} - \sqrt{e^{-b} \frac{b^x}{x!}} \right)^2 \left( \frac{2}{3} \sqrt{\frac{e^{-a} \frac{a^x}{x!}}{e^{-b} \frac{b^x}{x!}}} + \frac{1}{3} \right) \\ &= \sum_{x=0}^{\infty} \frac{(e^{-\frac{a}{2}} a^{\frac{x}{2}} - e^{-\frac{b}{2}} b^{\frac{x}{2}})^2}{x!} \left( \frac{2}{3} \sqrt{\frac{e^{b-a} a^x}{b^x}} + \frac{1}{3} \right) \\ &\leq (a-b)^2 e^{-L} \sum_{x=0}^{\infty} \frac{(U^{\frac{x}{2}} + xU^{\frac{x}{2}-1})^2}{x!} \left( \frac{2}{3} \sqrt{\frac{e^{U-L} U^x}{L^x}} + \frac{1}{3} \right) \\ &\leq (a-b)^2 e^{\frac{U-3L}{2}} \sum_{x=0}^{\infty} \frac{(U^{\frac{x}{2}} + xU^{\frac{x}{2}-1})^2}{x!} \left( \frac{U}{L} \right)^{\frac{x}{2}} \lesssim (a-b)^2, \end{aligned}$$

where the first inequality follows from the inequality  $|e^{-\frac{a}{2}} a^{\frac{x}{2}} - e^{-\frac{b}{2}} b^{\frac{x}{2}}| \leq |a-b| e^{-\frac{L}{2}} (U^{\frac{x}{2}} + xU^{\frac{x}{2}-1})$  for all  $a, b \in [L, U]$ . This implies that  $\frac{1}{n} \sum_{i=1}^n H_{*,i}(p_{\theta_1,i}, p_{\theta_2,i})^2 \lesssim \int (\theta_1 - \theta_2)^2 d\mathbb{P}_n^z$  for all link functions  $\theta_1$  and  $\theta_2$ , where  $\mathbb{P}_n^z = n^{-1} \sum_{i=1}^n \delta_{z_i}$  denotes the empirical distribution of  $z_1, z_2, \dots, z_n$ . So one can use the  $L_2(\mathbb{P}_n^z)$ -metric to produce the partition  $\{B_1, \dots, B_{K_n}\}$  of the space of link functions. By Theorem 2.7.5 of [10] we know that  $\log K_n \lesssim \varepsilon_n^{-1}$ . Letting  $\varepsilon_n^{-1} = n\varepsilon_n^2$  we obtain  $\varepsilon_n = n^{-1/3}$ , and hence by Theorem 8 the posterior based on uniform distributions converges almost surely at the rate  $\varepsilon_n = n^{-1/3}$  with respect to the metric  $d_n^0$ , which is the minimax rate for this model. The

in-probability convergence rate  $n^{-1/3}$  for the posterior based on discrete distributions has been obtained in Section 7.1.1 of Ghosal and van der Vaart [5].

It is worth pointing out that in this example the supnorm  $\|p_{\theta_{1,i}}/p_{\theta_{2,i}}\|_{\infty}$  may not be finite. Therefore, the approach on determination of prior concentration rates by means of  $H(p_{\theta_{1,i}}, p_{\theta_{2,i}}) \|p_{\theta_{1,i}}/p_{\theta_{2,i}}\|_{\infty}$  in Ghosal, Ghosh and van der Vaart [4] fails to be applied in this case, but the modified Hellinger distance  $H_*(p_{\theta_{1,i}}, p_{\theta_{2,i}})$  works well. A similar argument holds even for the infinite-dimensional normal model.

## References

- [1] Barron, A., Schervish, M. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536-561.
- [2] Cox, D.D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21**, 903-923.
- [3] Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite dimensional parameters. *Ann. Statist.* **27**, 1119-1140.
- [4] Ghosal, S., Ghosh, J.K. and van der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28**, 500-531.
- [5] Ghosal, S. and van der Vaart, A.W. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35**, 192-223.
- [6] Le Cam, L. (1986). *Asymptotic Methods In Statistical Decision Theory*. Springer, New York.
- [7] Pollard, D. (1990). *Empirical Processes: Theory And Applications*. IMS, Hayward, CA.
- [8] Scricciolo, C. (2006). Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34**, 2897-2920.
- [9] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29**, 687-714.
- [10] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer, New York.

- [11] Walker, S. (2003). On sufficient conditions for Bayesian consistency. *Biometrika*, **90**, 2, 482-488.
- [12] Walker, S. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028-2043.
- [13] Walker, S. G. and Hjort, N.L. (2001). On Bayesian consistency. *J.R. Statist. Soc., B* **63**, 811-821.
- [14] Walker, S., Lijoi, A. and Prunster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Statist.* **35**, 738-746.
- [15] Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *J. R. Statist. Soc., B* **62**, 159-180.
- [16] Xing, Y. (2008). Convergence rates of nonparametric posterior distributions. available at [www.arxiv.org](http://www.arxiv.org): 0804.2733.
- [17] Xing, Y. (2008). On adaptive Bayesian inference. *Electronic J. Statist.* **2**, 848-862.
- [18] Xing, Y. and Ranneby, B. (2008). Sufficient conditions for Bayesian consistency. To appear in *J. Statist. Plann. Inference*.
- [19] Xing, Y. and Ranneby, B. (2008). On Bayesian consistency. Research Report 2008:9, Centre of Biostochastics, Swedish University of Agricultural Sciences.
- [20] Zhao, L.H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28**, 532-552.

## Appendix

*Proof of Proposition 1.* Given  $\delta > 1$ , by the definition of the Hausdorff  $\alpha$ -constant and Assumption 1, there exist pairwise disjoint subsets  $B_1, B_2, \dots, B_{N_n}$  of  $\Theta_1$  such that (1)  $\cup_{k=1}^{N_n} B_k = \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}$ ; (2) each  $B_k$  is contained in some ball of  $e_n$ -radius not exceeding  $\varepsilon$ ; (3)  $\sum_{k=1}^{N_n} \Pi_n(B_k)^\alpha \leq \delta C(\varepsilon, \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}, \alpha, e_n)$ ; (4) there exist test functions  $\phi_k$  such that  $P_{\theta_0}^{(n)} \phi_k \leq e^{-Kn\varepsilon^2}$  and  $P_\theta^{(n)} \phi_k \geq 1 - e^{-Kn\varepsilon^2}$  for all  $\theta$  in  $B_k$ . Then by the inequality  $(x + y)^\alpha \leq x^\alpha + y^\alpha$  for all  $x, y \geq 0$ , we get

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq \sum_{k=1}^{N_n} P_{\theta_0}^{(n)} \left( \int_{B_k} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq \sum_{k=1}^{N_n} P_{\theta_0}^{(n)} \left\{ \phi_k^{1-\alpha} \left( \int_{B_k} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \right\} \\
& \quad + \sum_{k=1}^{N_n} P_{\theta_0}^{(n)} \left\{ (1 - \phi_k)^{1-\alpha} \left( \int_{B_k} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \right\} \\
& := L_1 + L_2.
\end{aligned}$$

It turns out from Hölder's inequality and Fubini's theorem that

$$\begin{aligned}
L_1 & \leq \sum_{k=1}^{N_n} (P_{\theta_0}^{(n)} \phi_k)^{1-\alpha} \left( P_{\theta_0}^{(n)} \int_{B_k} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq e^{-(1-\alpha)Kn\varepsilon^2} \sum_{k=1}^{N_n} \left( \int_{B_k} P_{\theta_0}^{(n)} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& = e^{-(1-\alpha)Kn\varepsilon^2} \sum_{k=1}^{N_n} \Pi_n(B_k)^\alpha \\
& \leq \delta e^{-(1-\alpha)Kn\varepsilon^2} C(\varepsilon, \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}, \alpha, e_n).
\end{aligned}$$

To estimate  $L_2$ , we deal with  $1/2 \leq \alpha < 1$  and  $0 < \alpha < 1/2$  separately. In the case of  $1/2 \leq \alpha < 1$  we have  $0 \leq (2\alpha - 1)/\alpha < 1$  and by Hölder's inequality,

$$\begin{aligned}
L_2 &= \sum_{k=1}^{N_n} P_{\theta_0}^{(n)} \left\{ (1 - \phi_k)^{1-\alpha} \left( \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{1-\alpha} \right. \\
&\quad \left. \cdot \left( \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{2\alpha-1} \right\} \\
&\leq \sum_{k=1}^{N_n} \left\{ P_{\theta_0}^{(n)} \left( (1 - \phi_k) \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right) \right\}^{1-\alpha} \\
&\quad \cdot \left\{ P_{\theta_0}^{(n)} \left( \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{\frac{2\alpha-1}{\alpha}} \right\}^{\alpha} \\
&\leq \sum_{k=1}^{N_n} \left( \int_{B_k} P_{\theta}^{(n)}(1 - \phi_k) \Pi_n(d\theta) \right)^{1-\alpha} \left( P_{\theta_0}^{(n)} \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{2\alpha-1} \\
&\leq \sum_{k=1}^{N_n} e^{-(1-\alpha)Kn\varepsilon^2} \Pi_n(B_k)^{1-\alpha} \Pi_n(B_k)^{2\alpha-1} \\
&= e^{-(1-\alpha)Kn\varepsilon^2} \sum_{k=1}^{N_n} \Pi_n(B_k)^{\alpha} \\
&\leq \delta e^{-(1-\alpha)Kn\varepsilon^2} C(\varepsilon, \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}, \alpha, e_n).
\end{aligned}$$

In the case of  $0 < \alpha < 1/2$  we have  $0 \leq (1 - \phi_k)^{1-\alpha} \leq (1 - \phi_k)^{\alpha} \leq 1$  and hence by Hölder's inequality,

$$\begin{aligned}
L_2 &\leq \sum_{k=1}^{N_n} P_{\theta_0}^{(n)} \left\{ \left( (1 - \phi_k) \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{\alpha} \right\} \\
&\leq \sum_{k=1}^{N_n} \left\{ P_{\theta_0}^{(n)} \left( (1 - \phi_k) \int_{B_k} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right) \right\}^{\alpha} \\
&= \sum_{k=1}^{N_n} \left( \int_{B_k} P_{\theta}^{(n)}(1 - \phi_k) \Pi_n(d\theta) \right)^{\alpha} \leq e^{-\alpha Kn\varepsilon^2} \sum_{k=1}^{N_n} \Pi_n(B_k)^{\alpha} \\
&\leq \delta e^{-\alpha Kn\varepsilon^2} C(\varepsilon, \{\theta \in \Theta_1 : d_n(\theta, \theta_0) > \varepsilon\}, \alpha, e_n).
\end{aligned}$$

Thus for any  $0 < \alpha < 1$  we have obtained the required inequality for  $K_1 = 2\delta$  and  $K_2 = \alpha K$  if  $0 < \alpha < 1/2$  and  $K_2 = (1 - \alpha)K$  if  $1/2 \leq \alpha < 1$ . Finally, letting  $\delta \searrow 1$ , we conclude the proof of Proposition 1.  $\square$

*Proof of Proposition 2.* Take nonempty disjoint subsets  $B_j$ ,  $j = 1, 2, \dots, N$ , of  $\Theta$  such that  $\sum_{j=1}^N \Pi(B_j)^\alpha \leq 2C(\delta\varepsilon, \{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon\}, \alpha, d_n^1)$ ,  $\cup_{j=1}^N B_j = \{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon\}$  and  $d_n^1$ -diameters of all  $B_j$  do not exceed  $2\delta\varepsilon$ . Then we have

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq P_{\theta_0}^{(n)} \sum_{j=1}^N \left( \int_{B_j} R_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& = \sum_{j=1}^N \Pi_n(B_j)^\alpha P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^\alpha \\
& \leq 2C(\delta\varepsilon, \{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon\}, \alpha, d_n^1) \max_{1 \leq j \leq N} P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^\alpha,
\end{aligned}$$

where  $I_j(X^{(n)}) = \Pi_n(B_j)^{-1} \int_{B_j} p_\theta^{(n)}(X^{(n)}) \Pi_n(d\theta)$  is the integral mean of the likelihood  $p_\theta^{(n)}(X^{(n)})$  and hence is a density function. With a slight abuse of notation we also let  $I_j$  stand for the corresponding parameter of this integral means. Take  $\theta_j \in B_j$  for each  $j$ . By Jensen's inequality for  $d_n^1(\cdot, \theta_j)^s$  we have  $d_n^1(I_j, \theta_j) \leq 2\delta\varepsilon$  and thus  $d_n^1(I_j, \theta_0) \geq d_n^1(\theta_j, \theta_0) - d_n^1(I_j, \theta_j) \geq (1 - 2\delta)\varepsilon$ . Take an nonnegative integer  $m$  with  $\frac{\alpha}{1-\alpha} \leq 2^m < \frac{2\alpha}{1-\alpha}$ . From Hölder's inequality it turns out that for each  $j$ ,

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^\alpha \\
& = P_{\theta_0}^{(n)} \left( \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2}} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2}} \right) \\
& \leq \left( P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2} \cdot \frac{2}{2-\alpha}} \right)^{\frac{2-\alpha}{2}} \left( P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2} \cdot \frac{2}{\alpha}} \right)^{\frac{\alpha}{2}} \\
& = \left( P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2-\alpha}} \right)^{\frac{2-\alpha}{2}},
\end{aligned}$$

which, by repeating the above procedure  $m - 1$  more times, does not exceed

$$\begin{aligned}
& \left( P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}} \right)^{\frac{2^m(1-\alpha)+\alpha}{2^m}} \\
& \leq \left( P_{\theta_0}^{(n)} \left( \frac{I_j(X^{(n)})}{p_{\theta_0}^{(n)}(X^{(n)})} \right)^{\frac{1}{2}} \right)^{\frac{\alpha}{2^{m-1}}} \\
& = \left( \int \sqrt{I_j(X^{(n)}) p_{\theta_0}^{(n)}(X^{(n)})} \mu(dX^{(n)}) \right)^{\frac{\alpha}{2^{m-1}}} \\
& = \left( 1 - \frac{H(I_j(X^{(n)}) p_{\theta_0}^{(n)}(X^{(n)}))^2}{2} \right)^{\frac{\alpha}{2^{m-1}}} \leq e^{-2^{-m} \alpha n d_n^1(I_j \theta_0)^2} \\
& \leq e^{-2^{-m} (1-2\delta)^2 \alpha n \varepsilon^2} \leq e^{-\frac{1}{2}(1-\alpha)(1-2\delta)^2 n \varepsilon^2},
\end{aligned}$$

which completes the proof of Proposition 2.  $\square$

*Proof of Proposition 3.* Denote  $S = \{\theta \in \Theta_1 : d_n^1(\theta, \theta_0) > \varepsilon\}$ . Assume first  $0 < \beta \leq 1/2$ . By Hölder's inequality and the inequality  $1 - x \leq e^{-x}$ , we have

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{\beta} \Pi_n(d\theta) \right)^{\alpha} \\
& \leq P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{\beta \cdot \frac{1}{2\beta}} \Pi_n(d\theta) \right)^{2\beta\alpha} \Pi_n(S)^{(1-2\beta)\alpha} \\
& \leq \left( P_{\theta_0}^{(n)} \int_S R_{\theta}^{(n)}(X^{(n)})^{\frac{1}{2}} \Pi_n(d\theta) \right)^{2\beta\alpha} \Pi_n(S)^{(1-2\beta)\alpha} \\
& = \left( \int_S P_{\theta_0}^{(n)} R_{\theta}^{(n)}(X^{(n)})^{\frac{1}{2}} \Pi_n(d\theta) \right)^{2\beta\alpha} \Pi_n(S)^{(1-2\beta)\alpha} \\
& \leq \left( \int_S e^{-\frac{1}{2}n\varepsilon^2} \Pi_n(d\theta) \right)^{2\beta\alpha} \Pi_n(S)^{(1-2\beta)\alpha} = e^{-\beta\alpha n \varepsilon^2} \Pi_n(S)^{\alpha},
\end{aligned}$$

which gives the required inequality when  $0 < \beta \leq 1/2$ . If  $1/2 < \beta < 1$  we take  $p = \frac{1}{2-2\beta}$  and  $q = \frac{1}{2\beta-1}$ . It then follows from Hölder's inequality that

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{\beta} \Pi_n(d\theta) \right)^{\alpha} \\
& \leq P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{\frac{1}{2p} \cdot p} \Pi_n(d\theta) \right)^{\frac{\alpha}{p}} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{(\beta - \frac{1}{2p}) \cdot q} \Pi_n(d\theta) \right)^{\frac{\alpha}{q}} \\
& \leq \left( P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)})^{\frac{1}{2}} \Pi_n(d\theta) \right)^{\alpha} \right)^{\frac{1}{p}} \left( P_{\theta_0}^{(n)} \left( \int_S R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{\alpha} \right)^{\frac{1}{q}} \\
& \leq \left( P_{\theta_0}^{(n)} \int_S R_{\theta}^{(n)}(X^{(n)})^{\frac{1}{2}} \Pi_n(d\theta) \right)^{\frac{\alpha}{p}} \left( P_{\theta_0}^{(n)} \int_S R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{\frac{\alpha}{q}} \\
& \leq \left( \int_S e^{-\frac{1}{2}n\varepsilon^2} \Pi_n(d\theta) \right)^{\frac{\alpha}{p}} \Pi_n(S)^{\frac{\alpha}{q}} = e^{-(1-\beta)\alpha n\varepsilon^2} \Pi_n(S)^{\alpha}.
\end{aligned}$$

The proof of Proposition 3 is complete.  $\square$

To prove Theorem 1 we need two simple lemmas.

**Lemma 1.** *Let  $\varepsilon > 0$  and  $c > 0$ . Then the inequality*

$$P_{\theta_0}^{(n)} \left( \int_{\Theta} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \leq e^{-n\varepsilon^2(3+2c)} \Pi_n(W_n(\theta_0, \varepsilon)) \right) \leq e^{-n\varepsilon^2 c}$$

holds for all  $n$ .

*Proof.* Without loss of generality, we may assume that  $\Pi_n(W_n(\theta_0, \varepsilon)) > 0$ . From Jensen's inequality and Chebyshev's inequality it follows that

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\Theta} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \leq e^{-n\varepsilon^2(3+2c)} \Pi_n(W_n(\theta_0, \varepsilon)) \right) \\
& \leq P_{\theta_0}^{(n)} \left( e^{n\varepsilon^2(\frac{3}{2}+c)} \leq \left( \frac{1}{\Pi_n(W_n(\theta_0, \varepsilon))} \int_{W_n(\theta_0, \varepsilon)} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{-\frac{1}{2}} \right) \\
& \leq P_{\theta_0}^{(n)} \left( e^{n\varepsilon^2(\frac{3}{2}+c)} \leq \frac{1}{\Pi_n(W_n(\theta_0, \varepsilon))} \int_{W_n(\theta_0, \varepsilon)} R_{\theta}^{(n)}(X^{(n)})^{-\frac{1}{2}} \Pi_n(d\theta) \right) \\
& \leq \frac{\int_{W_n(\theta_0, \varepsilon)} P_{\theta_0}^{(n)} R_{\theta}^{(n)}(X^{(n)})^{-\frac{1}{2}} \Pi_n(d\theta)}{e^{n\varepsilon^2(\frac{3}{2}+c)} \Pi_n(W_n(\theta_0, \varepsilon))},
\end{aligned}$$



where

$$\begin{aligned}
& P_{\theta_0}^{(n)} R_{\theta}^{(n)} (X^{(n)})^{-\frac{1}{2}} \\
&= 1 + \int \left( \sqrt{p_{\theta_0}^{(n)}(X^{(n)})} - \sqrt{p_{\theta}^{(n)}(X^{(n)})} \right)^2 \frac{\sqrt{p_{\theta_0}^{(n)}(X^{(n)})}}{\sqrt{p_{\theta}^{(n)}(X^{(n)})}} \mu(dx) \\
&\quad + \int \left( p_{\theta_0}^{(n)}(X^{(n)}) - \sqrt{p_{\theta}^{(n)}(X^{(n)})} p_{\theta_0}^{(n)}(X^{(n)}) \right) \mu(dx) \\
&= 1 + \int \left( \sqrt{p_{\theta_0}^{(n)}(X^{(n)})} - \sqrt{p_{\theta}^{(n)}(X^{(n)})} \right)^2 \frac{\sqrt{p_{\theta_0}^{(n)}(X^{(n)})}}{\sqrt{p_{\theta}^{(n)}(X^{(n)})}} \mu(dx) \\
&\quad + \frac{1}{2} \int \left( \sqrt{p_{\theta_0}^{(n)}(X^{(n)})} - \sqrt{p_{\theta}^{(n)}(X^{(n)})} \right)^2 \mu(dx) \\
&= 1 + \frac{3}{2} H_*(p_{\theta_0}^{(n)}, p_{\theta}^{(n)})^2 \leq e^{\frac{3}{2} n \varepsilon^2},
\end{aligned}$$

which implies the required inequality and the proof of Lemma 1 is complete.  $\square$

**Lemma 2.** *Under Assumption 2, the inequality*

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1 : d_n(\theta, \theta_0) \geq r\varepsilon} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^{\alpha} \\
&\leq K_1 \sum_{j=[r-1]}^{\infty} e^{-K_2 n j^2 \varepsilon^2} C(j\varepsilon, \{\theta \in \Theta_1 : j\varepsilon < d_n(\theta, \theta_0) \leq 2j\varepsilon\}, \alpha, e_n)^{K_3}
\end{aligned}$$

holds for all  $r \geq 2$ ,  $\varepsilon > 0$ ,  $\Theta_1 \subset \Theta$  and for all  $n$  large enough.

*Proof.* Note that  $\{\theta \in \Theta_1 : d_n(\theta, \theta_0) \geq r\varepsilon\} \subset \{\theta \in \Theta_1 : d_n(\theta, \theta_0) \geq [r]\varepsilon\} = \cup_{j=[r-1]}^{\infty} \{\theta \in \Theta_1 : j\varepsilon < d_n(\theta, \theta_0) \leq 2j\varepsilon\} := \cup_{j=[r-1]}^{\infty} \Theta_{1,j}$ . Using the inequality  $(x+y)^{\alpha} \leq x^{\alpha} + y^{\alpha}$  for all  $x, y \geq 0$  and Assumption 2 for  $\Theta_1 = \Theta_{1,j}$  we obtain

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_1: d_n(\theta, \theta_0) \geq r\varepsilon} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq \sum_{j=[r-1]}^{\infty} P_{\theta_0}^{(n)} \left( \int_{\Theta_{1,j}} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha \\
& \leq K_1 \sum_{j=[r-1]}^{\infty} e^{-K_2 n j^2 \varepsilon^2} C(j\varepsilon, \{\theta \in \Theta_{1,j} : d_n(\theta, \theta_0) > j\varepsilon\}, \alpha, e_n)^{K_3} \\
& = K_1 \sum_{j=[r-1]}^{\infty} e^{-K_2 n j^2 \varepsilon^2} C(j\varepsilon, \{\theta \in \Theta_1 : j\varepsilon < d_n(\theta, \theta_0) \leq 2j\varepsilon\}, \alpha, e_n)^{K_3}.
\end{aligned}$$

The proof of Lemma 2 is complete.  $\square$

*Proof of Theorem 1.* Take a constant  $c > 1/c_0$ . Then  $e^{-n\varepsilon_n^2 c} \leq e^{-cc_0 \log n} = 1/n^{cc_0}$  and hence  $\sum_{n=1}^{\infty} e^{-n\varepsilon_n^2 c} < \infty$ . By Lemma 1 and the first Borel-Cantelli lemma, we get that for almost all  $X^{(n)}$  the inequality

$$\int_{\Theta} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \geq e^{-n\varepsilon_n^2 (3+2c)} \Pi_n(W_n(\theta_0, \varepsilon_n))$$

holds for all large  $n$ . Thus, for any  $\delta > 0$  we have

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)}) \geq \delta \right) \\
& = P_{\theta_0}^{(n)} \left( \delta^{-\alpha} \Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)})^\alpha \geq 1 \right) \\
& \leq \delta^{-\alpha} P_{\theta_0}^{(n)} \left( \Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r\varepsilon_n | X^{(n)})^\alpha \right) \\
& \leq \delta^{-\alpha} e^{\alpha n \varepsilon_n^2 (3+2c)} \Pi_n(W_n(\theta_0, \varepsilon_n))^{-\alpha} \\
& \quad \cdot P_{\theta_0}^{(n)} \left( \int_{\theta \in \Theta_n: d_n(\theta, \theta_0) \geq r\varepsilon_n} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \right)^\alpha,
\end{aligned}$$

which, by Lemma 2 and the inequality (2), does not exceed

$$\begin{aligned}
& K_1 \delta^{-\alpha} e^{\alpha n \varepsilon_n^2 (3+2c)} \sum_{j=[r-1]}^{\infty} e^{-K_2 n j^2 \varepsilon_n^2 + c_1 j^2 n \varepsilon_n^2} \\
& \leq K_1 \delta^{-\alpha} e^{\alpha n \varepsilon_n^2 (3+2c)} \sum_{j=[r-1]}^{\infty} e^{(c_1 - K_2) j n \varepsilon_n^2} \\
& = \frac{K_1 e^{(c_1 - K_2)[r-1] n \varepsilon_n^2 + \alpha(3+2c)n \varepsilon_n^2}}{\delta^\alpha (1 - e^{(c_1 - K_2)n \varepsilon_n^2})} \leq \frac{K_1 n^{(c_1 - K_2)[r-1]c_0 + \alpha(3+2c)c_0}}{\delta^\alpha (1 - n^{(c_1 - K_2)c_0})} \\
& \leq 2K_1 \delta^{-\alpha} n^{(c_1 - K_2)[r-1]c_0 + \alpha(3+2c)c_0},
\end{aligned}$$

where the next last inequality holds for all large  $r$  and the last inequality holds for all large  $n$ . Since the last exponent is strictly less than  $-1$  for all large  $r$ , by the first Borel-Cantelli lemma we obtain that for almost all  $X^{(n)}$ ,

$$\Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \leq \delta$$

if  $n$  is large enough, which yields the first assertion.

To get the second assertion, choose a positive constant  $b$  with  $c_2 - \frac{b}{2} > \frac{1}{c_0}$ . We then follow the above proof, but take  $c = c_2 - \frac{b}{2}$  and  $\delta = e^{-bn\varepsilon_n^2}$  instead, and note that

$$\begin{aligned} & P_{\theta_0}^{(n)}\left(\Pi_n(\theta \in \Theta : d_n(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \geq e^{-bn\varepsilon_n^2}\right) \\ & \leq P_{\theta_0}^{(n)}\left(\Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \geq \frac{1}{2} e^{-bn\varepsilon_n^2}\right) \\ & \quad + P_{\theta_0}^{(n)}\left(\Pi_n(\theta \in \Theta \setminus \Theta_n : d_n(\theta, \theta_0) \geq r \varepsilon_n | X^{(n)}) \geq \frac{1}{2} e^{-bn\varepsilon_n^2}\right), \end{aligned}$$

where by Lemma 1 the second term on the right hand side is dominated by

$$\begin{aligned} & \frac{2e^{bn\varepsilon_n^2} e^{n\varepsilon_n^2(3+2c_2-b)}}{\Pi_n(W_n(\theta_0, \varepsilon_n))} P_{\theta_0}^{(n)} \int_{\theta \in \Theta \setminus \Theta_n : d_n(\theta, \theta_0) \geq r \varepsilon_n} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \\ & \leq \frac{2e^{n\varepsilon_n^2(3+2c_2)}}{\Pi_n(W_n(\theta_0, \varepsilon_n))} \int_{\Theta \setminus \Theta_n} P_{\theta_0}^{(n)} R_{\theta}^{(n)}(X^{(n)}) \Pi_n(d\theta) \\ & = \frac{2e^{n\varepsilon_n^2(3+2c_2)} \Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(W_n(\theta_0, \varepsilon_n))}. \end{aligned}$$

Then, using the same argument as the above, one can easily prove the second assertion and the proof of Theorem 1 is complete.  $\square$

Using the trivial inequality  $C(\delta\varepsilon_n, \Theta_n, \alpha, e_n) \leq C(\varepsilon_n, \Theta_n, \alpha, e_n)$  for  $\delta \geq 1$ , it is easy to see that Theorem 2 follows from Theorem 1. The proof of Theorem 3 is only a slight modification of the proof of Theorem 1 except that we need to apply Lemma 10 in Ghosal and van der Vaart [5]. The proof of Theorem 4 is completely similar to the proof of Theorem 1, but instead of an application of Lemma 1 one needs the following Lemma.

**Lemma 3.** *For independent observations  $(X_1, X_2, \dots, X_n)$  we have that the inequality*

$$P_{\theta_0}^{(n)}\left(\int_{\Theta} \left(\prod_{i=1}^n \frac{p_{\theta,i}(X_i)}{p_{\theta_0,i}(X_i)}\right)^{\beta} \Pi_n(d\theta) \leq e^{-n\varepsilon^2(3+2c)\beta} \Pi_n(\overline{W}_n(\theta_0, \varepsilon))\right) \leq e^{-n\varepsilon^2 c}$$

holds for all  $n$ ,  $\varepsilon > 0$ ,  $c > 0$  and  $0 < \beta \leq 1$ .

*Proof of Lemma 3.* Similar to the proof of Lemma 1 one can get that

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \int_{\Theta} \left( \prod_{i=1}^n \frac{p_{\theta,i}(X_i)}{p_{\theta_0,i}(X_i)} \right)^\beta \Pi_n(d\theta) \leq e^{-n\varepsilon^2(3+2c)\beta} \Pi_n(\overline{W}_n(\theta_0, \varepsilon)) \right) \\
& \leq P_{\theta_0}^{(n)} \left( e^{n\varepsilon^2(\frac{3}{2}+c)} \leq \left( \frac{\int_{\overline{W}_n(\theta_0, \varepsilon)} \left( \prod_{i=1}^n \frac{p_{\theta,i}(X_i)}{p_{\theta_0,i}(X_i)} \right)^\beta \Pi_n(d\theta)}{\Pi_n(\overline{W}_n(\theta_0, \varepsilon))} \right)^{-\frac{1}{2\beta}} \right) \\
& \leq P_{\theta_0}^{(n)} \left( e^{n\varepsilon^2(\frac{3}{2}+c)} \leq \frac{\int_{\overline{W}_n(\theta_0, \varepsilon)} \left( \prod_{i=1}^n \frac{p_{\theta,i}(X_i)}{p_{\theta_0,i}(X_i)} \right)^{-\frac{1}{2}} \Pi_n(d\theta)}{\Pi_n(\overline{W}_n(\theta_0, \varepsilon))} \right) \\
& \leq \frac{\int_{\overline{W}_n(\theta_0, \varepsilon)} \prod_{i=1}^n (1 + \frac{3}{2} H_{*,i}(p_{\theta_1,i}, p_{\theta_2,i})^2) \Pi_n(d\theta)}{e^{n\varepsilon^2(\frac{3}{2}+c)} \Pi_n(\overline{W}_n(\theta_0, \varepsilon))} \\
& \leq \frac{\int_{\overline{W}_n(\theta_0, \varepsilon)} e^{\frac{3}{2} \sum_{i=1}^n H_{*,i}(p_{\theta_1,i}, p_{\theta_2,i})^2} \Pi_n(d\theta)}{e^{n\varepsilon^2(\frac{3}{2}+c)} \Pi_n(\overline{W}_n(\theta_0, \varepsilon))} \\
& \leq e^{-n\varepsilon^2 c},
\end{aligned}$$

which concludes the proof.  $\square$

*Proof of Proposition 5.* It is no restriction to assume that  $n = 2k$  is an even number. Similar to the proof of Proposition 2 we get that the left side of the required inequality does not exceed  $2C(\delta\varepsilon, \{\theta \in \Theta_1 : d(\theta, \theta_0) > \varepsilon\}, \alpha, d)$  times

$$\begin{aligned}
& \max_{1 \leq j \leq N} P_{\theta_0}^{(n)} \left( \frac{1}{\Pi_n(B_j)} \int_{B_j} \frac{q_\theta(X_0) \prod_{i=1}^{2k} p_\theta(X_i | X_{i-1})}{q_{\theta_0}(X_0) \prod_{i=1}^{2k} p_{\theta_0}(X_i | X_{i-1})} \Pi_n(d\theta) \right)^\alpha \\
& = \max_{1 \leq j \leq N} P_{\theta_0}^{(n)} \left( \frac{\int_{B_j} q_\theta(X_0) \Pi_n(d\theta)}{q_{\theta_0}(X_0) \Pi_n(B_j)} \prod_{s=0}^{2k-1} \frac{I_{j,s}}{p_{\theta_0}(X_{s+1} | X_s)} \right)^\alpha = \\
& \max_{1 \leq j \leq N} P_{\theta_0}^{(n)} \left( \left( \frac{\int_{B_j} q_\theta(X_0) \Pi_n(d\theta)}{q_{\theta_0}(X_0) \Pi_n(B_j)} \prod_{t=1}^k \frac{I_{j,2t-1}}{p_{\theta_0}(X_{2t} | X_{2t-1})} \right)^\alpha \left( \prod_{t=0}^{k-1} \frac{I_{j,2t}}{p_{\theta_0}(X_{2t+1} | X_{2t})} \right)^\alpha \right) \\
& \leq \max_{1 \leq j \leq N} \left( P_{\theta_0}^{(n)} \left( \frac{\int_{B_j} q_\theta(X_0) \Pi_n(d\theta)}{q_{\theta_0}(X_0) \Pi_n(B_j)} \prod_{t=1}^k \frac{I_{j,2t-1}}{p_{\theta_0}(X_{2t} | X_{2t-1})} \right)^{2\alpha} \right)^{\frac{1}{2}}
\end{aligned}$$

$$\max_{1 \leq j \leq N} \left( P_{\theta_0}^{(n)} \left( \prod_{t=0}^{k-1} \frac{I_{j,2t}}{p_{\theta_0}(X_{2t+1}|X_{2t})} \right)^{2\alpha} \right)^{\frac{1}{2}} := \left( \max_{1 \leq j \leq N} A_{j,k} \right) \left( \max_{1 \leq j \leq N} B_{j,k} \right),$$

where the last inequality follows from Hölder's inequality, the set  $B_j$  is defined in a similar way as that of Proposition 2 and we have used the notations  $\prod_{i=1}^0 p_{\theta}(X_i|X_{i-1}) = 1$  and

$$I_{j,s} = \frac{\int_{B_j} q_{\theta}(X_0) \prod_{i=1}^{s+1} p_{\theta}(X_i|X_{i-1}) \Pi_n(d\theta)}{\int_{B_j} q_{\theta}(X_0) \prod_{i=1}^s p_{\theta}(X_i|X_{i-1}) \Pi_n(d\theta)}$$

for  $s = 0, 1, \dots, 2k-1$ . We also let  $I_{j,s}$  stand for the parameter of the corresponding integral means. Take  $\theta_j \in B_j$  for each  $j$ . From Jensen's inequality and the assumption  $a_0 r(X_s) \leq p_{\theta}(X_s|X_{s-1}) \leq a_1 r(X_s)$  it turns out that

$$\begin{aligned} d(I_{j,s}, \theta_j)^2 &= \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{I_{j,s}} - \sqrt{p_{\theta_j}(X_{s+1}|X_s)})^2 d\mu(X_{s+1}) d\nu(X_s) \\ &\leq \int_{B_j} \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{p_{\theta}(X_{s+1}|X_s)} - \sqrt{p_{\theta_j}(X_{s+1}|X_s)})^2 d\mu(X_{s+1}) \\ &\quad \frac{q_{\theta}(X_0) \prod_{i=1}^s p_{\theta}(X_i|X_{i-1})}{\int_{B_j} q_{\theta}(X_0) \prod_{i=1}^s p_{\theta}(X_i|X_{i-1}) \Pi_n(d\theta)} d\nu(X_s) \Pi_n(d\theta) \\ &\leq \frac{a_1}{a_0} \int_{B_j} \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{p_{\theta}(X_{s+1}|X_s)} - \sqrt{p_{\theta_j}(X_{s+1}|X_s)})^2 d\mu(X_{s+1}) d\nu(X_s) \\ &\quad \frac{q_{\theta}(X_0) \prod_{i=1}^{s-1} p_{\theta}(X_i|X_{i-1})}{\int_{B_j} q_{\theta}(X_0) \prod_{i=1}^{s-1} p_{\theta}(X_i|X_{i-1}) \Pi_n(d\theta)} \Pi_n(d\theta) \\ &= \frac{a_1}{a_0} \int_{B_j} d(\theta, \theta_j)^2 \frac{q_{\theta}(X_0) \prod_{i=1}^{s-1} p_{\theta}(X_i|X_{i-1})}{\int_{B_j} q_{\theta}(X_0) \prod_{i=1}^{s-1} p_{\theta}(X_i|X_{i-1}) \Pi_n(d\theta)} \Pi_n(d\theta) \\ &\leq \frac{4a_1 \delta^2 \varepsilon^2}{a_0} \end{aligned}$$

Thus,  $d(I_{j,s}, \theta_j) \leq \frac{2\sqrt{a_1} \delta \varepsilon}{\sqrt{a_0}}$  and  $d(I_{j,s}, \theta_0) \geq d(\theta_j, \theta_0) - d(I_{j,s}, \theta_j) \geq (1 - \frac{2\sqrt{a_1} \delta}{\sqrt{a_0}}) \varepsilon$ .

Write

$$\begin{aligned}
& A_{j,k}^2 \\
&= \int_{\mathfrak{X}^{2k-1}} \left( \int_{\mathfrak{X}} \left( \int_{\mathfrak{X}} \left( \frac{I_{j,2k-1}}{p_{\theta_0}(X_{2k}|X_{2k-1})} \right)^{2\alpha} d\mu(X_{2k}) \right) p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \right) \\
&\quad \cdot \left( \frac{\int_{B_j} q_{\theta}(X_0) \Pi_n(d\theta)}{q_{\theta_0}(X_0) \Pi_n(B_j)} \prod_{t=1}^{k-1} \frac{I_{j,2t-1}}{p_{\theta_0}(X_{2t}|X_{2t-1})} \right)^{2\alpha} \\
&\quad \cdot q_{\theta_0}(X_0) \prod_{s=0}^{2k-3} p_{\theta_0}(X_{s+1}|X_s) d\mu(X_0) d\mu(X_1) \dots d\mu(X_{2k-2}).
\end{aligned}$$

Take an nonnegative integer  $m$  with  $\frac{2\alpha}{1-2\alpha} \leq 2^m < \frac{4\alpha}{1-2\alpha}$ . Repeating the proof of Proposition 2 (applying the same procedure  $m+1$  times instead of  $m$  times) we get that

$$\begin{aligned}
& \int_{\mathfrak{X}} \left( \int_{\mathfrak{X}} \left( \frac{I_{j,2k-1}}{p_{\theta_0}(X_{2k}|X_{2k-1})} \right)^{2\alpha} d\mu(X_{2k}) \right) p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \\
&\leq \int_{\mathfrak{X}} \left( 1 - \frac{1}{2} \int_{\mathfrak{X}} (\sqrt{I_{j,2k-1}} - \sqrt{p_{\theta_0}(X_{2k}|X_{2k-1})})^2 d\mu(X_{2k}) \right)^{\frac{\alpha}{2^{m-1}}} \\
&\quad \cdot p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \\
&\leq \left[ 1 - \frac{1}{2} \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{I_{j,2k-1}} - \sqrt{p_{\theta_0}(X_{2k}|X_{2k-1})})^2 p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k}) d\mu(X_{2k-1}) \right]^{\frac{\alpha}{2^{m-1}}} \\
&\leq \left( 1 - \frac{a_0}{2} \int_{\mathfrak{X}} \int_{\mathfrak{X}} (\sqrt{I_{j,2k-1}} - \sqrt{p_{\theta_0}(X_{2k}|X_{2k-1})})^2 d\mu(X_{2k}) d\nu(X_{2k-1}) \right)^{\frac{1}{2}-\alpha} \\
&= \left( 1 - \frac{a_0 d(I_{j,2k-1}, \theta_0)^2}{2} \right)^{\frac{1}{2}-\alpha} \leq e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2} - \sqrt{a_1} \delta)^2 \varepsilon^2}.
\end{aligned}$$

Hence we have

$$\begin{aligned}
A_{j,k}^2 &\leq e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2} - \sqrt{a_1} \delta)^2 \varepsilon^2} \int_{\mathfrak{X}^{2k-1}} \left( \frac{\int_{B_j} q_{\theta}(X_0) \Pi_n(d\theta)}{q_{\theta_0}(X_0) \Pi_n(B_j)} \prod_{t=1}^{k-1} \frac{I_{j,2t-1}}{p_{\theta_0}(X_{2t}|X_{2t-1})} \right)^{2\alpha} \\
&\quad q_{\theta_0}(X_0) \prod_{s=0}^{2k-3} p_{\theta_0}(X_{s+1}|X_s) d\mu(X_0) d\mu(X_1) \dots d\mu(X_{2k-2}).
\end{aligned}$$

Repeating the same argument  $k - 1$  times one can get that

$$\begin{aligned}
A_{j,k}^2 &\leq e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2}-\sqrt{a_1}\delta)^2k\varepsilon^2} \int_{\mathfrak{X}} \left( \frac{\int_{B_j} q_\theta(X_0)\Pi_n(d\theta)}{q_{\theta_0}(X_0)\Pi_n(B_j)} \right)^{2\alpha} q_{\theta_0}(X_0) d\mu(X_0) \\
&\leq e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2}-\sqrt{a_1}\delta)^2k\varepsilon^2} \left( \int_{\mathfrak{X}} \frac{\int_{B_j} q_\theta(X_0)\Pi_n(d\theta)}{q_{\theta_0}(X_0)\Pi_n(B_j)} q_{\theta_0}(X_0) d\mu(X_0) \right)^{2\alpha} \\
&= e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2}-\sqrt{a_1}\delta)^2k\varepsilon^2}.
\end{aligned}$$

Similarly, we have

$$B_{j,k}^2 \leq e^{-(1-2\alpha)(\frac{\sqrt{a_0}}{2}-\sqrt{a_1}\delta)^2k\varepsilon^2}.$$

Hence we have proved the required inequality and the proof of Proposition 5 is complete.  $\square$

The proof of Theorem 6 is completely similar to that of Theorem 1 except that we apply Proposition 5 and the following lemma.

**Lemma 4.** *If there exists a constant  $a_1 \geq 1$  such that  $\int_A p_{\theta_0}(y|x)d\mu(y) \leq a_1 \int_A d\nu(y)$  for all  $x \in \mathfrak{X}$  and  $A \in \mathcal{A}$ , then the inequality*

$$\begin{aligned}
P_{\theta_0}^{(n)} \left( \int_{\Theta} \frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{i=1}^n \frac{p_\theta(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} \Pi_n(d\theta) \leq e^{-n\varepsilon^2(3a_1+4c)} \Pi_n(W_n^1(\theta_0, \varepsilon)) \right) \\
\leq e^{-n\varepsilon^2c}
\end{aligned}$$

holds for all  $n, \varepsilon > 0$  and  $c > 0$ .

*Proof of Lemma 4.* Similar to the proof of Lemma 1 we have that the left hand side of the required inequality does not exceed

$$\frac{\int_{W_n^1(\theta_0, \varepsilon)} P_{\theta_0}^{(n)} \left( \frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{i=1}^n \frac{p_\theta(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} \right)^{\frac{1}{4}} \Pi_n(d\theta)}{e^{n\varepsilon^2(\frac{3a_1}{4}+c)} \Pi_n(W_n^1(\theta_0, \varepsilon))}.$$

So it suffices to prove that  $P_{\theta_0}^{(n)} \left( \frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{i=1}^n \frac{p_\theta(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} \right)^{\frac{1}{4}} \leq e^{\frac{3a_1}{4}n\varepsilon^2}$  for all  $\theta \in W_n^1(\theta_0, \varepsilon)$ . We assume without loss of generality that  $n$  is an even number, say  $n = 2k$ . Write

$$\frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{i=1}^n \frac{p_\theta(X_i|X_{i-1})}{p_{\theta_0}(X_i|X_{i-1})} = \frac{q_\theta(X_0)}{q_{\theta_0}(X_0)} \prod_{j=1}^k \frac{p_\theta(X_{2j}|X_{2j-1})}{p_{\theta_0}(X_{2j}|X_{2j-1})} \prod_{j=1}^k \frac{p_\theta(X_{2j-1}|X_{2j-2})}{p_{\theta_0}(X_{2j-1}|X_{2j-2})}.$$

From Hölder's inequality it then turns out that

$$\begin{aligned}
& P_{\theta_0}^{(n)} \left( \frac{q_{\theta_0}(X_0)}{q_{\theta}(X_0)} \prod_{i=1}^n \frac{p_{\theta_0}(X_i|X_{i-1})}{p_{\theta}(X_i|X_{i-1})} \right)^{\frac{1}{4}} \\
& \leq \left[ P_{\theta_0}^{(n)} \left( \frac{q_{\theta_0}(X_0)}{q_{\theta}(X_0)} \prod_{j=1}^k \frac{p_{\theta_0}(X_{2j}|X_{2j-1})}{p_{\theta}(X_{2j}|X_{2j-1})} \right)^{\frac{1}{2}} \right]^{\frac{1}{2}} \left[ P_{\theta_0}^{(n)} \left( \prod_{j=1}^k \frac{p_{\theta_0}(X_{2j-1}|X_{2j-2})}{p_{\theta}(X_{2j-1}|X_{2j-2})} \right)^{\frac{1}{2}} \right]^{\frac{1}{2}} \\
& := A_k B_k.
\end{aligned}$$

Hence by Fubini's theorem we get that  $A_k^2$  is equal to

$$\begin{aligned}
& \int_{\mathfrak{X}^{2k+1}} \frac{q_{\theta_0}(X_0)^{\frac{3}{2}}}{q_{\theta}(X_0)^{\frac{1}{2}}} \prod_{j=1}^k \left( \frac{p_{\theta_0}(X_{2j}|X_{2j-1})^{\frac{3}{2}}}{p_{\theta}(X_{2j}|X_{2j-1})^{\frac{1}{2}}} p_{\theta_0}(X_{2j-1}|X_{2j-2}) \right) \\
& \quad d\mu(X_0) d\mu(X_1) \dots d\mu(X_{2k}) \\
& = \int_{\mathfrak{X}^{2k-1}} \left( \int_{\mathfrak{X}} \left( \int_{\mathfrak{X}} \frac{p_{\theta_0}(X_{2k}|X_{2k-1})^{\frac{3}{2}}}{p_{\theta}(X_{2k}|X_{2k-1})^{\frac{1}{2}}} d\mu(X_{2k}) \right) p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \right) \\
& \quad \frac{q_{\theta_0}(X_0)^{\frac{3}{2}}}{q_{\theta}(X_0)^{\frac{1}{2}}} \prod_{j=1}^{k-1} \frac{p_{\theta_0}(X_{2j}|X_{2j-1})^{\frac{3}{2}}}{p_{\theta}(X_{2j}|X_{2j-1})^{\frac{1}{2}}} p_{\theta_0}(X_{2j-1}|X_{2j-2}) d\mu(X_0) d\mu(X_1) \dots d\mu(X_{2k-2}),
\end{aligned}$$

where by the proof of Lemma 1 in Xing [16] we have

$$\begin{aligned}
& \int_{\mathfrak{X}} \left( \int_{\mathfrak{X}} \frac{p_{\theta_0}(X_{2k}|X_{2k-1})^{\frac{3}{2}}}{p_{\theta}(X_{2k}|X_{2k-1})^{\frac{1}{2}}} d\mu(X_{2k}) \right) p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \\
& = \int_{\mathfrak{X}} \left( 1 + \frac{3}{2} H_*(p_{\theta_0}(\cdot|X_{2k-1}), p_{\theta}(\cdot|X_{2k-1}))^2 \right) p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \\
& = 1 + \int_{\mathfrak{X}} \frac{3}{2} H_*(p_{\theta_0}(\cdot|X_{2k-1}), p_{\theta}(\cdot|X_{2k-1}))^2 p_{\theta_0}(X_{2k-1}|X_{2k-2}) d\mu(X_{2k-1}) \\
& \leq 1 + \int_{\mathfrak{X}} \frac{3a_1}{2} H_*(p_{\theta_0}(\cdot|X_{2k-1}), p_{\theta}(\cdot|X_{2k-1}))^2 d\nu(X_{2k-1}) \\
& = 1 + \frac{3a_1}{2} H_*(p_{\theta_0}, p_{\theta})^2 \leq e^{\frac{3a_1}{2} H_*(p_{\theta_0}, p_{\theta})^2}.
\end{aligned}$$

Thus, we have obtained that  $A_k \leq e^{\frac{3a_1}{4} H_*(p_{\theta_0}, p_{\theta})^2} A_{k-1}$ . Repeating the same



argument  $k - 1$  times and using  $a_1 \geq 1$  one can get

$$\begin{aligned}
A_k &\leq e^{\frac{3a_1}{4}kH_*(p_{\theta_0}, p_{\theta})^2} \left( \int_{\mathfrak{X}} \frac{q_{\theta_0}(X_0)^{\frac{3}{2}}}{q_{\theta}(X_0)^{\frac{1}{2}}} d\mu(X_0) \right)^{\frac{1}{2}} \\
&= e^{\frac{3a_1}{4}kH_*(p_{\theta_0}, p_{\theta})^2} \left( 1 + \frac{3}{2}H_*(q_{\theta_0}, q_{\theta})^2 \right)^{\frac{1}{2}} \\
&\leq e^{\frac{3}{4}H_*(q_{\theta_0}, q_{\theta})^2 + \frac{3a_1}{4}kH_*(p_{\theta_0}, p_{\theta})^2}.
\end{aligned}$$

Similarly, we can get that  $B_k \leq e^{\frac{3a_1}{4}kH_*(p_{\theta_0}, p_{\theta})^2}$ . Therefore  $A_k B_k \leq e^{\frac{3}{4}H_*(q_{\theta_0}, q_{\theta})^2 + \frac{3a_1}{4}nH_*(p_{\theta_0}, p_{\theta})^2} \leq e^{\frac{3a_1}{4}n\varepsilon^2}$  for all  $\theta \in W_n^1(\theta_0, \varepsilon)$ , and the proof of Lemma 4 is complete. □